DOCUMENT RESUME

ED 124 201 95 IR 003 595

AUTHOR	Borich, Gary D.; And Others
TITLE	The Analysis of Aptitude-Treatment Interactions: Computer Programs and Calculations.
INSTITUTION	Texas Univ., Austin. Research and Development Center for Teacher Education.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Mar 76
CONTRACT	NIE-C-74-0088
NOTE	106p.; Charts and parts of text may not reproduce clearly due to small type
AVAILABLE FROM	International Educational Services, P. O. Box A3650, Chicago, Illinois 60690 (ISBN-0-916378-03-9)
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	Annotated Bibliographies; *Computer Programs; Educational Psychology; Experimental Psychology; Instructional Systems; Interaction; Manuals; Psychological Evaluation; *Psychological Studies; Psychological Testing; Psychology; *Research Methodology; *Research Tools; *Statistical Analysis; Teacher Education
IDENTIFIERS	*Aptitude Treatment Interaction

ABSTRACT

The computer programs described in this manual were designed for the analysis of data resulting from psychological experiments which hypothesize an aptitude- or trait-treatment interaction (ATI). The programs were written in Fortran IV for use on most CDC 6400/6600 and IBM 360 computer systems. Research designs to which the programs are applicable represent paradigms which test for aptitude-by-treatment interactions under four conditions: when one continuous aptitude has been measured that is linearly related to the criterion (program ATILIN1); when two continuous attitudes have been measured that are linearly related to the criterion (program ATILIN2); when one continuous aptitude has been measured that is curvilinearly related to the criterion (program ATICURV); and when one aptitude has been measured that is linearly related to the criterion and group sizes are sufficiently large to permit a blocked analysis of variance on individuals with extreme scores in the aptitude distribution (program XGROUPS). Reference to technical and design issues is made in methodological notes which accompany each program. A short annotated bibliography appears at the end of the manual. (CH)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED124201

THE ANALYSIS OF APTITUDE-TREATMENT INTERACTIONS: COMPUTER PROGRAMS AND CALCULATIONS

Gary D. Borich, Robert C. Godbout
and Kenneth W. Wunderlich

Research and Development Center for Teacher Education

The University of Texas at Austin

Austin, Texas 78712

International Educational Services
P.O. Box A3650 Chicago, Illinois 60690
Cable Address: INEDSER

IR 003 595

Borich, Gary D., Robert D. Godbout & Kenneth W. Wunderlich.
The Analysis of Aptitude-Treatment Interactions:
Computer Programs and Calculations.
March 1976 Oasis Press, Austin, Texas
ISBN Number: 0-916378-03-9
Library of Congress Number: 76-5101

Table of Contents

	Introduction	ix, x
Chapter I.	Organization of the Programs	1
Chapter II.	System Control Cards	9
Chapter III.	Program ATILIN1	15
Chapter IV.	Program ATILIN2	43
Chapter V.	Program ATICURV	60
Chapter VI.	Program XGROUPS	77
Chapter VII.	Sample Data	94
Chapter VIII.	Annotated Bibliography	97

Dr. Borich is Director of the Instructional Systems Laboratory and Associate Professor of Educational Psychology at The University of Texas at Austin; Dr. Godbout is a Research Associate at the Research and Development Center for Teacher Education, The University of Texas at Austin; and Dr. Wunderlich is a member of the Advanced Studies in Education faculty at The University of Texas at San Antonio.

The authors wish to express their appreciation to David Malitz who edited and revised the present programs.

This work was funded, in part, by the National Institute of Education, Contract No. NIE-C-74-0088, The Evaluation of Teaching Project. The opinions expressed herein do not necessarily reflect the position or policies of the National Institute of Education and no official endorsement by that office should be inferred.

The programs described in this manual are available from:

International Educational Services
P. O. Box A3650 Chicago, Illinois 60690
Cable Address: INEDSER

INTRODUCTION

The computer programs described in this manual were designed for the analysis of data resulting from psychological experiments which hypothesize an aptitude- or trait-treatment interaction (ATI). The programs were written for four designs which either have appeared frequently in reported studies of ATI's or which have resulted from the authors' own research conducted at the Research and Development Center for Teacher Education, The University of Texas at Austin, and sponsored by grants from the U. S. Office of Education and the National Institute of Education.

The research designs to which the programs in this manual are applicable represent paradigms which test for aptitude-by-treatment interactions when (a) one continuous aptitude or trait has been measured that is linearly related to the criterion (program ATILIN1), (b) two continuous aptitudes or traits have been measured that are linearly related to the criterion (program ATILIN2), (c) one continuous aptitude or trait has been measured that is curvilinearly related to the criterion (program ATICURV), and (d) one aptitude or trait has been measured that is linearly related to the criterion and group sizes are sufficiently large to permit a treatment-by-blocks analysis of variance on individuals with extreme scores in the aptitude distribution (program XGROUPS). Program XGROUPS has been designed to increase the statistical power of an ATI analysis and thereby increase its sensitivity to detecting an aptitude-by-treatment interaction.

Each of the four programs to be described has been written for the case in which there are two treatments and either one or two aptitudes. These conditions need not be limiting, however, in that the programs can be easily modified to handle more complex conditions as the models for these are but simple extensions of those used to construct the present programs. Also, it has been the authors' experience in using the programs that, even when more than two treatments or aptitudes are present, the investigator will usually reduce the problem to a simpler condition for interpretative purposes. When there are more than two treatments, the investigator can conduct pairwise comparisons and, if there are more than two aptitudes, select pairs of aptitudes for which the aptitudes are minimally related to each other and maximally related to the criterion.

Reference to these and other technical issues are made in the *methodological notes* which accompany each of the programs described in this manual. The descriptions and illustrations contained within the methodological notes review and occasionally go beyond content provided by standard statistical texts which cover these pro-

cedures. The user is encouraged to study these notes to gain a full understanding of the purposes and statistical procedures employed by the programs. Such study will be helpful, if not prerequisite, to composing a full and accurate interpretation of results.

The user may also find helpful the annotated bibliography which appears at the conclusion of the manual and the following general articles about the programs themselves.

1. Program ATILIN1

One continuous aptitude linearly related to the criterion.

Borich, G. D. Interactions among group regressions: Testing homogeneity of group regressions and plotting regions of significance. *Educational and Psychological Measurement*, 1971, *31*, 251-253.

2. Program ATILIN2

Two continuous aptitudes linearly related to the criterion.

Borich, G. D., & Wunderlich, K. W. Johnson-Neyman revisited: Determining interactions among group regressions and plotting regions of significance in the case of two groups, two predictors and one criterion. *Educational and Psychological Measurement*, 1973, *33*, 155-159.

3. Program ATICURV

One continuous aptitude curvilinearly related to the criterion.

Wunderlich, K. W., & Borich, G. D. Determining interactions and regions of significance for curvilinear regressions. *Educational and Psychological Measurement*, 1973, *33*, 691-695.

4. Program XGROUPS

One aptitude linearly related to the criterion and large group sizes.

Borich, G. D., & Godbout, R. C. Extreme groups designs and the calculation of statistical power. *Educational and Psychological Measurement*, 1974, *34*, 663-675.

CHAPTER I

ORGANIZATION OF THE PROGRAMS

The programs described in this manual were written in Fortran IV versions that are appropriate for use on most CDC 6400/6600 and IBM 360 computer systems. With small modifications the programs can be made compatible with a variety of other computer systems, such as DEC-10 and UNIVAC. The user should consult his computation center to determine the applicability of the present programs.

The computer programs to be described follow a similar sequence of operations, this sequence being

1. read information from the title, parameter and format cards and read data (subroutine CCD),
2. delete missing and/or invalid data (subroutine AMISDAT), and
3. calculate summary statistics.

Programs ATILIN1, ATILIN2, and ATICURV:

4. test for the homogeneity of group regressions,
5. test for common intercepts (analysis of covariance),
6. define regions of significance,
7. determine point(s) or line at which regressions intersect, and
8. plot group regressions and regions of significance.

Program XGROUPS:

4. performs a treatment-by-blocks ANOVA on extreme scores, and
5. estimates statistical power for main and interaction effects.

Each of these operations is discussed briefly below.

Subroutine CCD

Subroutine CCD performs the first set of operations for all programs. This subroutine is used to input information from the control cards (the title control card, the parameter control card and the format control cards) which define each problem to be processed. Before returning control to ATILIN1, ATILIN2, ATICURV or XGROUPS, the subroutine prints the information contained on the three control

cards. At the start of the routine, the title control card is examined to determine whether this card is blank. If the title card is blank, a STOP statement terminates execution of the program. This characteristic allows the user to process multiple problems within a single job. When a problem is completed, subroutine CCD is called and the next card is read. If this card is blank then the last problem has been completed and the execution terminates. If this card is not blank, then it represents the title card for an additional problem and processing continues.

Multiple problem processing. When the user wishes to process multiple problems within a single job, data for all problems must follow the control cards for the first problem and precede the control cards for subsequent problems. During the first call to subroutine CCD, all data are read and then written in card image form on a scratch tape. The data appropriate to a given problem are obtained by reference to the data tape with the formats provided in the control cards for that problem. Such multiple problem processing is appropriate only when all problems deal with the same pair of treatment groups. Problems involving different pairs of treatment groups must be submitted as separate jobs.

Subroutine AMISDAT

Immediately after the call to subroutine CCD, subroutine AMISDAT is called for the purpose of deleting missing and/or invalid data. This routine is controlled by the missing data option (column 12 on the parameter card). When a 0 is punched in column 12 of the parameter card, subroutine AMISDAT will consider all data valid; when a 1 is punched in column 12, all blanks will be considered invalid; and when a 2 is punched in column 12, both blanks and 0's are considered invalid. In that the missing data subroutine precedes all other calculations, only data which meet the specifications coded in the missing data column of the parameter card are analyzed.

Calculate Summary Statistics

Following the missing data option sequence, each program calculates within-group summary statistics. For programs ATILIN1, ATILIN2 and ATICURV, summary statistics consist of group sizes, means, standard deviations, and correlations. Summary statistics for program XGROUPS consist of cell means for the two main effects (groups and

treatments) and for the groups-by-treatment interaction. Whenever standard deviations are calculated as summary statistics, the printed value will be a sample statistic and not an estimate of the parameter value. Therefore, for the table of summary statistics, standard deviations are calculated with the number of scores (N) in the denominator rather than the degrees of freedom (e.g., $N - 1$). Elsewhere, where unbiased estimates of the parameter value are required, the degrees of freedom are used in the denominator.

Homogeneity of Group Regressions Test

For programs ATILIN1, ATILIN2 and ATICURV, the fourth step in the sequence is a regression analysis test for the presence of an aptitude-by-treatment interaction. This homogeneity of group regressions test involves the null hypothesis that group regressions (criterion on aptitude) are parallel, or, equivalently, that group slopes are equal ($b_1 = b_2$). Rejection of this null hypothesis indicates the existence of an aptitude-by-treatment interaction. This regression method for determining aptitude-by-treatment interactions differs from the traditional factorial analysis of variance (ANOVA) methods in that an aptitude commonly dichotomized or trichotomized to fit the factorial structure of analysis of variance is used in the regression method as a continuous measure to describe as many different types of subjects as there are observed values of a particular aptitude. Cronbach and Snow (1973, p. 318) have shown that for the case in which there is a moderately strong interaction, the statistical power of a test which employs the aptitude as a continuous variable is superior to blocking the aptitude at the median, blocking at the 33rd and 67th percentiles or similar configurations that may be employed in a treatment-by-blocks ANOVA design. These authors have concluded that by employing the homogeneity of group regressions test rather than a median split, about one-third fewer cases are required to maintain the same level of power (see Borich, 1975).

Test of Common Intercepts (Analysis of Covariance)

Following the homogeneity of group regressions test, ATILIN1, ATILIN2 and ATICURV test for common intercepts. This test is commonly referred to as "analysis of covariance" and tests the significance of the difference between group regressions at the mean of the

aptitude variable. This test is not interpreted when the investigator rejects the null hypothesis for the homogeneity of group regressions test, i.e., finds an aptitude-by-treatment interaction. However, when group regressions are homogeneous, the investigator may test the hypothesis that the mean of Treatment 1 significantly differs from the mean of Treatment 2 when all subjects in Groups 1 and 2 score at the mean of the aptitude variable. The new, adjusted group means for this analysis are determined by inserting the aptitude mean into the regression equations for Group 1 and Group 2 and solving each for $\hat{Y}$, the adjusted group mean.

Points of Intersection

ATILIN1, ATILIN2 and ATICURV determine the point or points at which the regression lines (ATILIN1), planes (ATILIN2) or curves (ATICURV) intersect. In the case of the one-predictor, ATILIN1, model, there will be one point of intersection if the group regressions are not parallel. In the case of the two-aptitude, ATILIN2, model, a straight line defines the intersection of two nonparallel regression planes. This line is called the "line of nonsignificance" or the "line of no difference between regression planes." For ATICURV, there may be none, one, or two points of intersection. Points of intersection will be useful to the user in visualizing regions of significance which, if present, will always be defined to the right and/or left of the point(s) or line of intersection.

Regions of Significance

A region of significance describes a range of values of a predictor (aptitude) variable for which there are significant group differences on the criterion—that is, the separation of the regression lines is significantly different from 0. The difference between groups ($\hat{Y}_1 - \hat{Y}_2$) will be exactly 0 at a point or line of intersection and the group differences will be nonsignificant at values of the predictor variable which lie close to an intersection. Areas farther to the right and/or the left of an intersection in which group differences are significant define the regions of significance. There may be either one (to the right *or* left) or two (to the right *and* left) regions of significance about an intersection. The calculations of regions of significance that are performed by the programs in this manual follow or are extensions of formulas set out by Johnson and Neyman (1936). The user may be in-

terested in other, analogous procedures for calculating regions of significance which are reported in an article by Cahen and Linn (1971). This article is annotated in the bibliography that appears at the conclusion of this manual.

Plotting Regions of Significance

Programs ATILIN1, ATILIN2 and ATICURV employ a graphical output routine which constructs a scatter plot, draws the within-group regressions and delineates the regions of significance. The programs allow the user to choose a 12 × 10 inch paper plot, a microfiche plot for use as a photographic negative, or no plot. As the programs themselves do not calculate the number of cases which fall within the regions of significance, it is important that users use the plot option whenever possible to insure themselves that any regions of significance that may be defined are of practical importance.

Users will need to check with their computer installation to determine if a plotter is available. Users whose installations do not include a plotter must indicate the "no plot" option on the parameter control card, and these users may find it necessary to insert dummy subroutines named PLT, SYMBOL, NUMBER, LINE, PLOT, AXIS, SCALE, BGNPLT, and ENDPLT. The form of these dummy subroutines is as follows:

```
SUBROUTINE PLT (F1, F2, I)
RETURN
END
```

Appropriate subroutine name cards for the remaining plotting subroutines are:

```
SUBROUTINE BGNPLT (I1, F1, I2, I3)
SUBROUTINE ENDPLT
SUBROUTINE SYMBOL (F1, F2, F3, I1, F4, I2)
SUBROUTINE LINE (F1, F2, I1, I2, I3, I4)
SUBROUTINE SCALE (F1, F2, I1, I2)
SUBROUTINE AXIS (F1, F2, I1, I2, F3, F4, F5, F6).
```

The graphical output routine included in the present programs is appropriate to the CalComp plotting package (California Computer Products, Inc., Anaheim, California). It will be necessary to alter the

plotting routine within the present programs for use with other graphical output packages.

The graphical output routine included in the present programs requires the following subroutines:

PLT:	converts all pen (beam) movements from inches to device commands and creates an appropriate file of these commands.
SYMBOL:	draws any sequence of characters and symbols. The entry point SYMSET can be used to redefine any symbol. SYMBOL calls only PLT.
NUMBER:	draws the fixed decimal equivalent of the internal floating point number. NUMBER calls SYMBOL which calls PLT.
LINE:	plots a series of data points defined by X and Y, connecting the points with straight lines if requested. It may call SYMBOL as well as PLT.
SCALE:	examines a data array to determine optimum starting value in a scaling factor for use by AXIS and LINE in converting data units to plotter page dimensions. It is one of the subroutines that does not call any other routine.
AXIS:	draws an axis line with the appropriate scale annotation and title and calls SYMBOL and NUMBER as well as PLT.
BGNPLT:	begins plot.
ENDPLT:	ends plot.

Treatment by Blocks—ANOVA

This analysis is performed only by program XGROUPS. XGROUPS classifies an aptitude into high and low categories, each containing equal numbers of extreme cases. The analysis yields mean squares for treatments, levels (high vs. low categories), treatment by levels and

residual error. The *F*-ratio for treatment by levels is the test for an aptitude-treatment interaction.

Estimate Statistical Power

These calculations also are made only by program XGROUPS. XGROUPS calculates the statistical power $(1 - \beta)$ for the main effects (levels and treatments) and interaction (levels by treatments) for the treatment-by-blocks ANOVA. Power is calculated with function POWER, which provides power estimates corresponding to Cohen's (1969) tabled values for differing values of (1) degrees of freedom in the denominator of the *F*-ratio, (2) cell frequency and (3) the *F*-ratio required for significance.

Summary

The sequence of steps and output for the programs in this manual is summarized below.

Sequence of Steps and Output	ATIINT	ATIINT2	ATICURV	XGROUPS
Calls input subroutine CCD	X	X	X	X
Calls missing data subroutine AMISDAT	X	X	X	X
Calculates summary statistics	X	X	X	X
Tests homogeneity of group regressions	X	X	X	
Tests intercepts (ANCOVA)	X	X	X	
Defines point(s)/line of intersection	X	X	X	
Plots regions	X	X	X	
Performs treatment-by-blocks ANOVA				X
Estimates statistical power				X

References

Borich, G. D. A methodological note on trait-treatment interactions. *Psychological Reports*, 1975, *36*, 407-412.

Cahen, L. S., & Linn, R. L. Regions of significant criterion differences in aptitude-treatment-interaction research. *American Educational Research Journal*, 1971, *8*(3), 521-530.

Cohen, J. *Statistical power analysis for the behavioral sciences*. New York: Academic Press, 1969.

Cronbach, L. J., & Snow, R. E. Aptitudes and instructional methods. Stanford: Stanford University, College of Education, 1973 (mimeo).

Johnson, P.O., & Neyman, J. Tests of certain linear hypotheses and their applications to some educational problems. *Statistical Research Memoirs*, 1936, *1*, 57-63.

CHAPTER II

SYSTEM CONTROL CARDS

The outline below presents the sequence and description of the system control cards required for submission of the programs to The University of Texas at Austin Computation Center. Other users will need to adjust this sequence in accord with the requirements of their particular facility. Figure 1 (p.14) reiterates the control card sequence.

1. User identification card
2. Password card
3. Job card
4. BSTAT Macro card
5. 7/8/9 card
6. Program buffer card
7. Call program card
8. 7/8/9 card
9. Program control cards and data
10. End-of-file (EOF) card

1. *User Identification Card*

Col 1-7	user number assigned by Computation Center
Col 8	comma
Col 9-28	user's name (as many columns as needed)
Col 29	period

2. *Password Card*

Col 1-3	password (initially assigned by Computation Center)
Col 4-13	"=PASSWORD."

3. *Job Card*

Col 1-3	"JOB"
Col 4	comma
Col 5-10	"TM=010" (time limit in seconds)
Col 11	comma
Col 12-17	"PR=100" (print limit in pages)
Col 18	period

These limits may be omitted (with the preceding comma) if the value to be used is less than:

8 seconds of time

17 printed pages.

The example limits may be increased as needed. If the default limits are accepted, the JOB card may be omitted.

4. *BSTAT Macro Card*

Col 1-23	"EXECPF(2350,STAT,BSTAT)"

5. *7/8/9*

This card contains the numbers 7,8,9 punched in column one. This 7/8/9 card separates the preceding system control cards from the subsequent FORTRAN program cards (cards 6 and 7).

6. *Program Buffer Card*

This card begins in Col. 7 with the word "PROGRAM" followed by the word "SELECT" (starting in Col. 15). The buffers required for the desired program are then listed in parentheses (starting in Col. 22). The number and type of buffers will be different for each program selected. Therefore, the following table should be consulted in completing this card.

Program	Buffers Required
ATILIN1	(INPUT,OUTPUT,PLOT,TAPE7=PLOT,TAPE15)
ATILIN2	(INPUT,OUTPUT,PLOT,TAPE15)
ATICURV	(INPUT,OUTPUT,PLOT,TAPE15)
XGROUPS	(INPUT,OUTPUT,TAPE15)

Col 7-13 "PROGRAM"

Col 15-20 "SELECT"

Col 22-72 buffers

Example:
"PROGRAM SELECT (INPUT,OUTPUT,PLOT,TAPE7 = PLOT,TAPE15)"

7. *Call Program Card*

This card also begins in Col. 7 and is used to call the program desired.

Col 7-10 "CALL"

Col 12-18 name of program

Col 20 "$"

Col 22-24 "END"

Example: "CALL ATILIN1 $ END"

8. *7/8/9*

This card contains the numbers 7, 8, and 9 punched in column one. This 7/8/9 card separates the preceding FORTRAN program cards from the subsequent program control cards.

9. *Program Control Cards and Data*

The program is controlled with the following cards which are specific to each program. Chapters III, IV, V and VI of this manual will explain the preparation of these program control cards.

Alphanumeric title card

Parameter card(s)

F-mode variable format card for Group 1 indicating locations on data cards of relevant aptitude(s) and criterion

(Group 1 data, aptitude(s) first and criterion second)

F-mode variable format card for Group 2 indicating locations on data cards of relevant aptitude(s) and criterion

(Group 2 data, aptitude(s) first and criterion second)

A separate set of program control cards is required for each problem to be processed. Data for all problems must be included with the program control cards for the first problem. Group 1 data for *all* problems follow the Group 1 format control card for the first problem. Group 2 data for *all* problems follow the Group 2 format control card for the first problem. Program control cards for additional problems follow the Group 2 data. A blank card is inserted after the last set of program control cards. The sequence of program control cards and data is presented in Figure 1. *Note* that all problems submitted within a single job must concern the same pair of treatment groups. Problems involving different pairs of treatment groups must be submitted within separate jobs.

Data organization. Data should be arranged by subject within treatment group with all of the data for Subject 1 preceding the data for Subject 2, etc. Within a subject's data, a predictor (aptit

must precede any criterion variable included in a problem with that predictor variable. The predictor variable(s) and criterion variable appropriate to a specific problem are determined by the format cards included in the program control cards for that problem.

16. EOF Card

The numbers 6, 7, 8 and 9 are punched in column one. The EOF card indicates the end of a job.

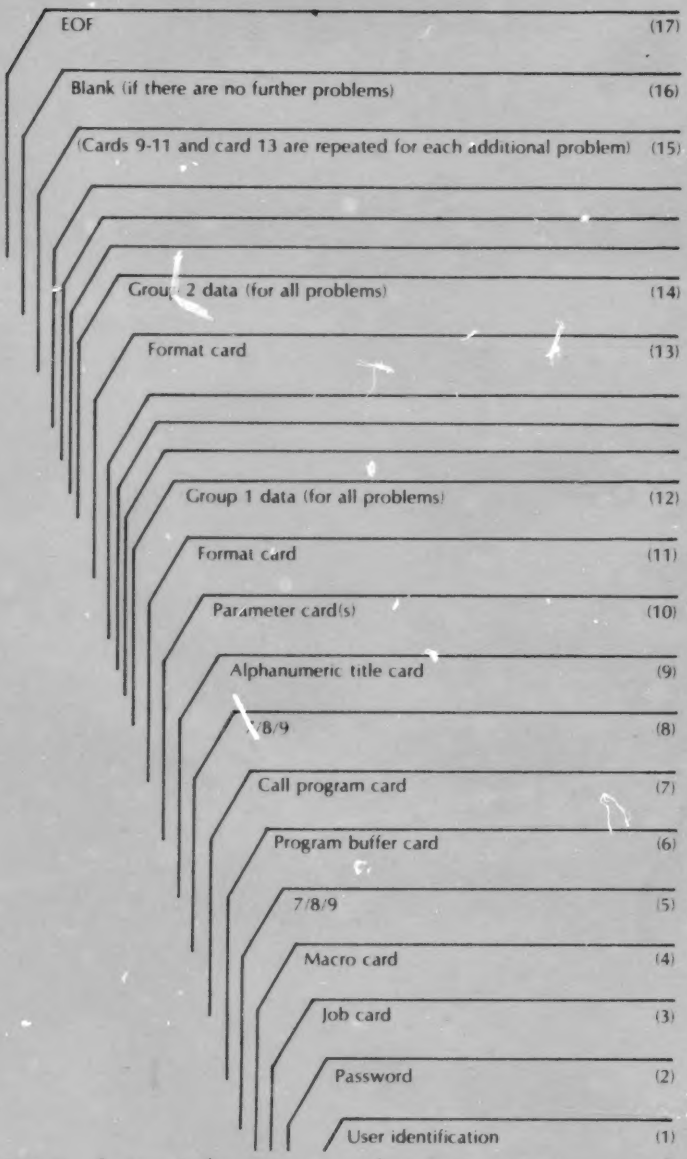

EOF (17)
Blank (if there are no further problems) (16)
(Cards 9-11 and card 13 are repeated for each additional problem) (15)
Group 2 data (for all problems) (14)
Format card (13)
Group 1 data (for all problems) (12)
Format card (11)
Parameter card(s) (10)
Alphanumeric title card (9)
7/8/9 (8)
Call program card (7)
Program buffer card (6)
7/8/9 (5)
Macro card (4)
Job card (3)
Password (2)
User identification (1)

Figure 1. *System and program control card arrangement.*

CHAPTER III

ATILIN1

Program Description

This program tests homogeneity of group regressions and defines regions of significance for the case in which there are two treatment groups and one continuous aptitude or trait that is linearly related to a criterion. Program outputs (1) table of summary statistics — group sizes, means, standard deviations, and correlations between aptitude and criterion; (2) regression equations (*Y*-intercepts and regression coefficients) for each group; (3) the aptitude value at which group regressions intersect; (4) *F*-value, degrees of freedom and probability for the homogeneity of group regressions test; (5) *F*-value, degrees of freedom and probability for the test of common intercepts (analysis of covariance); and (6) aptitude values for which treatment groups are significantly different (regions of significance). A flow chart for program ATILIN1 is presented in Figure 2.

Program Input

Card 1 alphanumeric title card Col 1-80

Card 2 parameter card

Col 1-5	*N* for Group 1 (maximum = 200)
Col 6-10	*N* for Group 2 (maximum = 200)
Col 12	missing data option 0 = all data valid 1 = blanks are invalid 2 = blanks and zeroes are invalid
Col 14	output option 0 = plot 1 = film 2 = printed output only
Col 16	option for table of predictor and criterion scores listed by subject within treatments. If this option is taken, ID

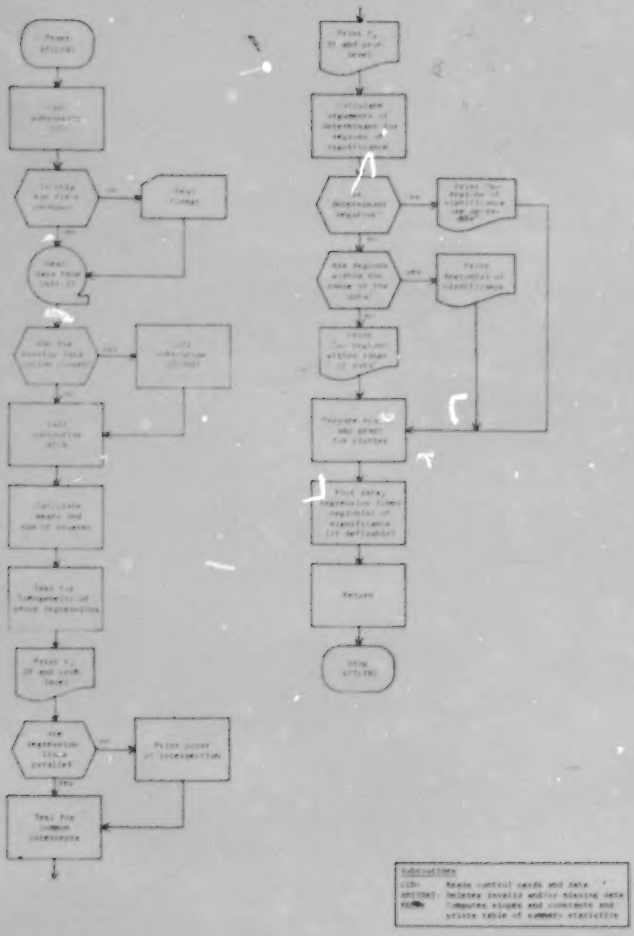


Figure 2. Program ATILIN1 flow chart.

codes will be read according to format and will be printed out (A5) along with corresponding predictor and criterion scores for subjects in each treatment group. Subjects with missing data will not be listed in this table.
0 = no table
1 = list ID codes and scores (begin format cards with A mode field)

Col 18 no. of cards per subject in Group 1
Col 20 no. of cards per subject in Group 2
Col 25-34 alpha level for regions of significance

Card 3 format for Group 1 Col 1-80
followed by Group 1 data

Card 4 format for Group 2 Col 1-80
followed by Group 2 data

Card 5 blank (after last problem)
for multiple problems repeat cards 1-4, omitting data

Data cards should contain subject ID codes (if desired), the aptitude score, and then the criterion score. If Col 16 on the parameter card is 0, then formats must specify two F-mode fields—the first field for the aptitude and the second for the criterion. If Col 16 on the parameter card is 1, then formats must specify an initial A-mode field (A5 or less) for the ID code and then the two F-mode fields.

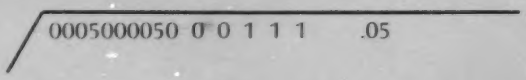
Example Problem

Data for this problem will be the first predictor and the criterion given as sample data in Chapter VII (p.95) of this manual. Program control cards for this example problem are as follows.

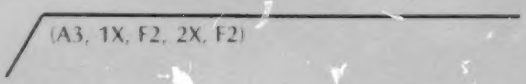
1. Alphanumeric title card

```
Example problem for ATILIN1
```

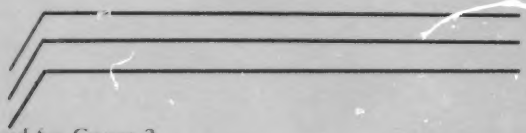
2. Parameter card



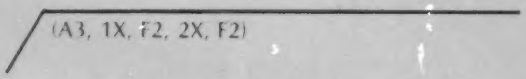
3. Format card for Group 1



4. Group 1 data

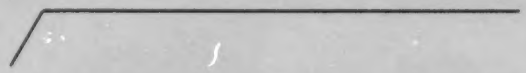


5. Format card for Group 2

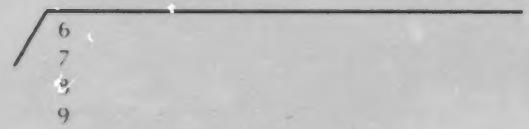


6. Group 2 data

7. Blank card (after last problem)



8. EOF



Printed output for this example problem is given in Figures 3 and 4. Plotted output is given in Figure 5.

TABLE OF PREDICTOR AND CRITERION SCORES LISTED BY SUBJECTS WITHIN TREATMENTS

TREATMENT 1

ID	PREDICTOR	CRITERION
001	10.000	5.000
002	15.000	15.000
003	20.000	15.000
004	25.000	20.000
005	30.000	20.000
006	20.000	25.000
007	35.000	25.000
008	40.000	25.000
009	30.000	30.000
010	40.000	30.000
011	50.000	30.000
012	25.000	35.000
013	30.000	35.000
014	45.000	35.000
015	55.000	35.000
016	35.000	40.000
017	40.000	40.000
018	50.000	40.000
019	35.000	45.000
020	55.000	45.000
021	60.000	45.000
022	65.000	45.000
023	40.000	50.000
024	45.000	50.000
025	50.000	50.000
026	60.000	50.000
027	65.000	50.000
028	70.000	50.000
029	50.000	55.000
030	50.000	60.000
031	55.000	60.000
032	60.000	60.000
033	65.000	60.000
034	70.000	60.000
035	75.000	60.000
036	60.000	65.000
037	70.000	65.000
038	80.000	65.000
039	60.000	70.000
040	65.000	70.000
041	70.000	70.000
042	75.000	70.000
043	85.000	70.000
044	70.000	75.000
045	80.000	75.000
046	75.000	80.000
047	85.000	80.000
048	90.000	80.000
049	85.000	85.000
050	90.000	85.000

Figure 3. *Printed output for ATILIN1 example problem—table of predictor and criterion scores.*

TREATMENT 2

ID	PREDICTOR	CRITERION
051	0.000	70.000
052	5.000	65.000
053	5.000	75.000
054	10.000	75.000
055	15.000	60.000
056	15.000	70.000
057	20.000	75.000
058	25.000	55.000
059	25.000	70.000
060	30.000	60.000
061	30.000	65.000
062	30.000	75.000
063	35.000	50.000
064	35.000	55.000
065	40.000	60.000
066	40.000	70.000
067	40.000	75.000
068	45.000	55.000
069	45.000	65.000
070	50.000	45.000
071	50.000	65.000
072	55.000	40.000
073	55.000	50.000
074	55.000	55.000
075	55.000	70.000
076	60.000	55.000
077	65.000	30.000
078	65.000	40.000
079	65.000	65.000
080	70.000	25.000
081	70.000	35.000
082	70.000	45.000
083	70.000	55.000
084	70.000	60.000
085	75.000	20.000
086	75.000	30.000
087	75.000	40.000
088	75.000	50.000
089	80.000	15.000
090	80.000	45.000
091	80.000	55.000
092	85.000	10.000
093	85.000	15.000
094	85.000	25.000
095	85.000	35.000
096	85.000	45.000
097	90.000	0.000
098	90.000	10.000
099	90.000	20.000
100	90.000	30.000

Figure 3 (continued). *Printed output for ATILIN1 example problem—table of predictor and criterion scores.*

```
        ATILIN1 SUMMARY STATISTICS

STATISTICS          GROUP 1          GROUP 2

N                        52               50

MEAN
  CRITERION         50.0000          48.5000
  PREDICTOR         54.1000          54.8000

SIGMA
  CRITERION         20.1742          19.9812
  PREDICTOR         20.6807          26.1526

CORRELATION           .9048            -.7900


THE REGRESSION EQUATION FOR GROUP 1 IS  Y =   2.2487  +     .8826  X

THE REGRESSION EQUATION FOR GROUP 2 IS  Y =  81.5742  +    -.6035  X


THE REGRESSION LINES INTERSECT AT THE PREDICTOR VALUE OF   53.3750


THE F-VALUE FOR THE TEST OF HOMOGENEITY OF GROUP REGRESSIONS IS  249.1984 WITH 1 AND   96
DEGREES OF FREEDOM WHICH HAS A PROBABILITY OF .0000


THE F-VALUE FOR THE TEST OF COMMON INTERCEPTS, ASSUMING HOMOGENEITY OF GROUP REGRESSIONS
IS     .1315 WITH 1 AND    97 DEGREES OF FREEDOM WHICH HAS A PROBABILITY OF .7185


LEFT REGION OF SIGNIFICANCE IS BOUNDED BY:    0.0000   50.4869
WHERE    0.0000  IS THE MINIMUM OBSERVED APTITUDE VALUE

RIGHT REGION OF SIGNIFICANCE IS BOUNDED BY:  56.2320   90.0000
WHERE   90.0000  IS THE MAXIMUM OBSERVED APTITUDE VALUE


END OF JOB.
```

Figure 4. *Printed output for ATILIN1 example problem—summary statistics and significance tests.*

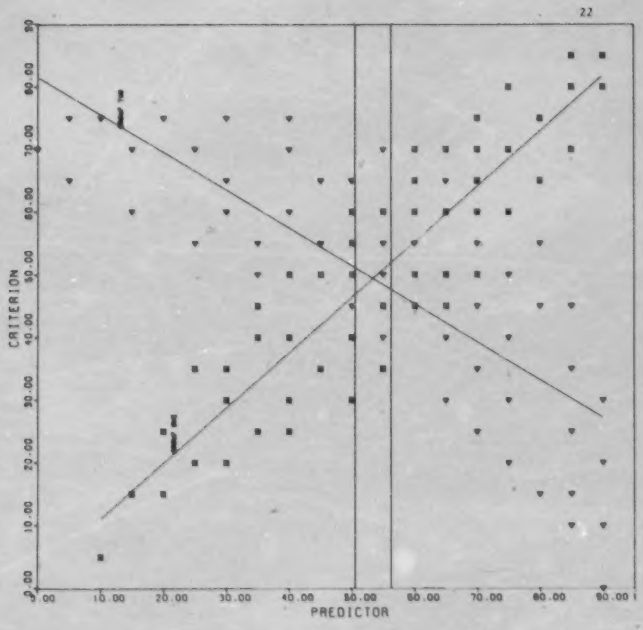


Figure 5. *Plotted output for ATILIN1 example problem.*

Methodological Notes*

ATIILIN1

1. *Homogeneity of Group Regressions Test*

To test the hypothesis that the regressions for two groups are parallel (i.e., the slopes are equal) ATIILIN1 constructs a standard linear prediction model of the form

$$Y_i = b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + b_4X_{4i} + e_i; \qquad [1]$$

$i = 1 \ldots, N$ (N being the total number of subjects in both groups);

where Y_i is the criterion; b_1 is the regression coefficient for the first group membership vector, X_{1i} (scored 1 if S_i is in Group 1, scored 0 if not); b_2, the regression coefficient of a second group membership vector, X_{2i} (scored 1 if S_i is in Group 2, scored 0 if not); b_3, the regression coefficient of the product of X_{1i} and the aptitude vector (this product symbolized as X_{3i}); and b_4, the regression coefficient of the product of X_{2i} and the aptitude vector (this product symbolized as X_{4i}).

The residual sum of squares (Σe_i^2) has degrees of freedom given by the number of Ss minus the number of linearly independent parameters. Therefore, we have $N - 4$ *df* or, for more than two treatment groups, $N - 2k$ *df*, where k equals the number of treatment groups.

To test that $b_3 = b_4$, i.e., that the regressions are homogeneous, the data are fitted to a second more restricted model which represents observations within each treatment group about regression lines with a common slope. This restricted model is of the form

---

*For explanatory purposes, methodological notes may reflect alternative but mathematically identical procedures to those actually employed by the programs.

$$Y_i = b_1X_{1i} + b_2X_{2i} + b_3(X_{3i} + X_{4i}) + f_i \qquad [2]$$

or alternatively

$$Y_i = b_1X_{1i} + b_2X_{2i} + b_3A_i + f_i$$

where A_i is the vector of aptitude scores. For the residual sum of squares (Σf_i^2) we have $N - 3$ *df* or, for more than two treatment groups, $N - k - 1$ *df*, where k equals the number of treatment groups. Since the restricted model [2] combines predictor variables treated separately in the starting model [1], Σf_i^2 is expected to be greater than Σe_i^2. These sums of squares can be equal if the null hypothesis is true (i.e., if group regessions are homogeneous), but Σf_i^2 cannot be less than Σe_i^2.

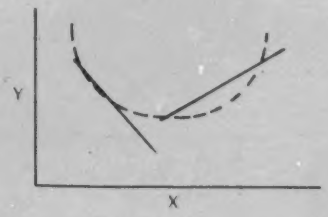
To test for homogeneous slopes, an hypothesis sum of squares is formed, given by $SS_{hyp} = \Sigma f_i^2 - \Sigma e_i^2$ with $(N - 3) - (N - 4) = 1$ *df* or, for models with k treatment groups, $k - 1$ *df*. The *F*-test* for homogeneous slopes is then given by

$$F(k - 1, N - 2k) = \frac{SS_{hyp} / (k - 1)}{\Sigma e_i^2 / (N - 2k)}$$

*This *F*-test illustrates the general method for significance testing within the context of multiple regression analysis. Significance testing proceeds according to the following steps. First, a starting or full model is written. Second, a restriction relevant to the hypothesis of interest is written. Third, a restricted model is formed by incorporating this restriction into the full model. Fourth, the hypothesis sum of squares (SS_{hyp}) is formed by subtracting the full model error sum of squares (SS_{full}) from the restricted model error sum of squares (SS_{rstrd}). The degrees of freedom for SS_{hyp} equal the degrees of freedom for SS_{rstrd} minus the degrees of freedom for SS_{full}. Fifth, the *F*-test for the hypothesis of interest is given by

$$F(df_{hyp}, df_{full}) = \frac{SS_{hyp} / df_{hyp}}{SS_{full} / df_{full}}.$$

A caution should be noted in the case where the distributions of aptitude scores for the two groups tend to have restricted overlap. In such an instance, heterogeneous linear regressions for two groups may simply reflect a *common* curvilinear regression as in the following diagram.



In such a case, the homogeneity of group regressions test may be significant but the conclusion that group regressions actually differ would be erroneous. The user should study the plotted output carefully to avoid such incorrect conclusions.

2. Test of Common Intercepts (Analysis of Covariance)

To test the hypothesis that two treatment groups are not significantly different when subjects score at the mean of the aptitude, ATILIN1 constructs a full model of the form

$$Y_i = b_1 X_{1i} + b_2 X_{2i} + b_3 X_{3i} + e_i$$

where Y_i is the criterion; b_1, the regression coefficient for the first group membership vector, X_{1i} (scored 1 if S_i is in Group 1, scored 0 if not); and b_2, the regression coefficient for the second group membership vector, X_{2i} (scored 1 if S_i is in Group 2, scored 0 if not); and b_3, the regression coefficient for the aptitude variable, X_{3i}. Note that this model is identical to [2]. The degrees of freedom for the residual sum of squares (Σe_i^2) for this model are $N - 3$ or, for the case of k treatment groups, $N - k - 1$.

To test for significant intercept differences, ATILIN1 constructs a second, more restricted model of the form

$$Y_i = a + b_3 X_{3i} + f_i$$

where a is the Y-intercept and $df = N - 2$. This restricted model results from placing the restriction $b_1 = b_2$ on the full model. An hypothesis sum of squares ($\Sigma f_i^2 - \Sigma e_i^2$) and F-ratio are constructed in the usual manner, using the error sums of squares from the full and restricted models.

3. *Point of Intersection (Test of Ordinality)*

An aptitude-treatment interaction is said to be *ordinal* when group regressions (criterion on aptitude) fail to intersect within the observed range of aptitude values and *disordinal* when group regressions do intersect within the observed range of aptitude values. Ordinality may be noted from a plot of the group regression lines or from determining the point of intersection as given by

$$X_0 = \frac{a_2 - a_1}{b_1 - b_2}$$

where a_1 and b_1 are the intercept and slope for the criterion on aptitude regression within Group 1 and a_2 and b_2 are corresponding values within Group 2. If X_0 falls within the range of observed aptitude values, then the interaction is disordinal; otherwise, the interaction is ordinal.

4. *Regions of Significance**

A region of significance describes a range of aptitude variable values for which there are significant group differences on the criterion—i.e., the distance between the regression lines is significantly different from 0. To identify regions of significance, the Johnson-Neyman technique (Johnson & Neyman, 1936; Johnson & Jackson, 1959) is employed. All values of the aptitude variable, for which there are significant differences between group regressions ($\hat{Y}_1 - \hat{Y}_2$), are in-

*To avoid confusion the reader should note that regions of significance as referred to in this manual are mathematically defined and, therefore, may not necessarily fall within the range of data.

cluded in regions of significance. When the difference, $\hat{Y}_1 - \hat{Y}_2$, for a single aptitude variable value is of interest, a *t*-statistic with $N_1 + N_2 - 4$ *df* can be computed as follows:

$$t = \frac{D}{\sqrt{S_D^2}}$$

with $D = \hat{Y}_1 - \hat{Y}_2$ and

$$S_D^2 = \frac{\left[\sum_{i=1}^{2} \left(Cyy_i - \frac{C^2xy_i}{Cxx_i}\right)\right]\left[\frac{N}{N_1N_2} + \sum_{i=1}^{2} \left(\frac{X' - \bar{X}_i^2}{Cxx_i}\right)\right]}{N_1 + N_2 - 4}$$

where *i* refers to treatment group; N_1 and N_2 are the numbers of subjects in the two treatments; $N = N_1 + N_2$; $Cxx = \Sigma X^2 - (\Sigma X)^2/N$ (sum of squares for *X*, the aptitude variable); $Cyy = \Sigma Y^2 - (\Sigma Y)^2/N$ (sum of squares for *Y*, the criterion variable); $Cxy = \Sigma XY - (\Sigma X)(\Sigma Y)/N$ (sum of cross products); X' is the aptitude value at which we are testing the distance between regression lines; and $\bar{X}$ is the mean of the aptitude variable.

Bounding aptitude variable values for regions of significance are obtained by solving the equation

$$X = \frac{-B \pm \sqrt{B^2 - AC}}{A}$$

where *A*, *B* and *C* represent terms provided by Walker and Lev (1953, p. 401). The equation for bounding values gives two (real) solutions when $B^2 - AC$ is greater than 0, one real solution when $B^2 - AC$ is exactly equal to 0, and no real solutions when $B^2 - AC$ is less than 0. A bounding value separates a region of significance from a region of nonsignificance. Consider an obtained bounding value, x. Does x represent the upper bound of a region of significance falling below it, or the lower bound of a region of significance falling above it? The point of intersection can be used to determine which alternative reflects the true state of affairs. At the point of intersection, the pre-

dicted criterion difference between the groups is 0. Such a zero difference can never be contained within a region of significant differences. That is, the point of intersection always lies within a region of nonsignificance. If the point of intersection falls in the region below a bounding value, then that bounding value is the lower bound for a region of significance falling above it. If the point of intersection falls in the region above the bounding value, then that bounding value is the upper bound for a region of significance falling below it.

Consider the location of regions of significance when no bounding values exist. In this case, a single region exists—that region extending from $-\infty$ to $+\infty$ on the aptitude variable. Given that the group regressions are not exactly parallel, then this region contains an intersection and it is a region of nonsignificance. As a general rule, a lack of bounding values indicates no regions of significance.*

The location of regions of significance when a single bounding value is obtained deserves little attention. A single bounding value represents a highly improbable case—the term $B^2 - AC$ in the bounding value equation is *exactly* equal to 0. The improbability of this situation is matched by a peculiar region of significance. When a single bounding value exists, then that single point is the entire region of significance with regions of nonsignificance falling both above and below it.

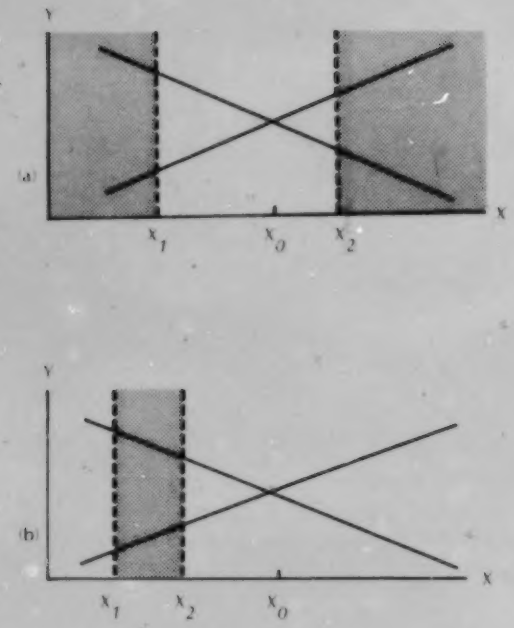
---

*An exception to this rule occurs when the group regressions are *exactly* parallel. Parallel regressions produce no intersection, and the single region may be either a region of significance or a region of nonsignificance. This exception is of little or no importance for two reasons. First, exactly parallel group regressions are highly improbable, and second, the Johnson-Neyman technique is not intended for use when group regressions are homogeneous. When regressions are homogeneous, simple analysis of covariance provides an adequate analysis strategy.

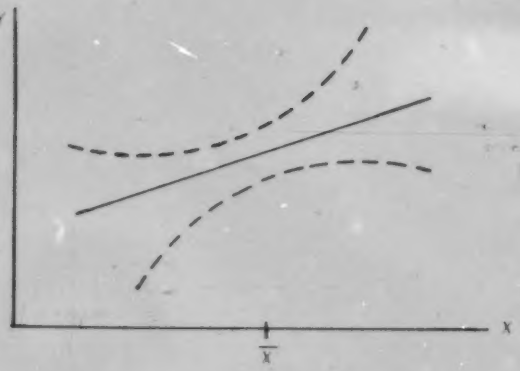
When two bounding values are obtained, then either one or two regions of significance exist. Let x_L represent the lower of two bounding values, and x_H the higher of two bounding values. If the intersection falls between x_L and x_H, then two regions of significance exist—one region extending from $-\infty$ to x_L, and the other region extending from x_H to $+\infty$. If the intersection falls below x_L or above x_H, then a single region of significance exists with x_L as its lower bound and x_H as its upper bound. Figures 6a and 6b illustrate one- and two-region cases.

The occurrence of a single region of significance may at first seem counterintuitive. The region of significance in Figure 6b has x_L as a lower bound. It may appear that this region should continue below x_1, since the distance beween the two regression lines increases. However, the test of significance involves not only the distance between the regression lines, but also the standard error of this distance—i.e., $t = D/S_D$. Given that D is increasing, it is not necessary that t also be increasing. If D and S_D are both increasing, and if S_D has the higher rate of increase, then t will be decreasing. Such a situation occurs in the one-region case.

Consideration of confidence intervals around the group regression lines may be helpful in clarifying the single-region case. The following diagram represents confidence intervals for Y constructed about the regression of Y on X.



Figures 6a and 6b. *Cases involving one region (b) and two regions (a) of significance. Shaded areas reflect regions of significance. X_0 indicates the point of intersection while X_1 and X_2 indicate obtained bounding values for regions of significance. Two regions occur when X_1 and X_2 straddle X_0 but one region occurs if both X_1 and X_2 fall to one side of X_0.*



The straight (solid) line represents the regression line, while the hyperbola (broken lines) depicts the confidence bounds. Note that the confidence limits are narrowest at the mean of X and expand as one moves away from $\overline{X}$ in both directions. Figure 6c presents the single-region case with confidence boundaries included. This figure demonstrates the plausibility of the single-region case.

Statistical texts (e.g., Johnson & Jackson, 1959; Kerlinger & Pedhazur, 1973; Walker & Lev, 1953) which deal with the Johnson-Neyman technique have not considered the single-region case. While such an oversight may lead to confusion on the part of the researcher who does obtain a single region of significance, the importance of such an oversight is mitigated by two points. First, the single-region case does not occur with great frequency and, second, the single-region case may actually be somewhat of an anomaly.

To illustrate these two points, consider the aptitude-treatment interaction results reported by Borich, Godbout, Peck, Kash & Poynor (1974). These authors report 107 significant aptitude-treatment interactions, as evidenced by heterogeneous group regression slopes significant at the .10 level. Regions of significance analyses were computed for each of these 107 interactions. The significance level employed in these latter analyses was .05. Of the 107 significant interactions, only 13 yielded a single region of significance, while 61 produced two regions and 33 produced no regions. Of special in-

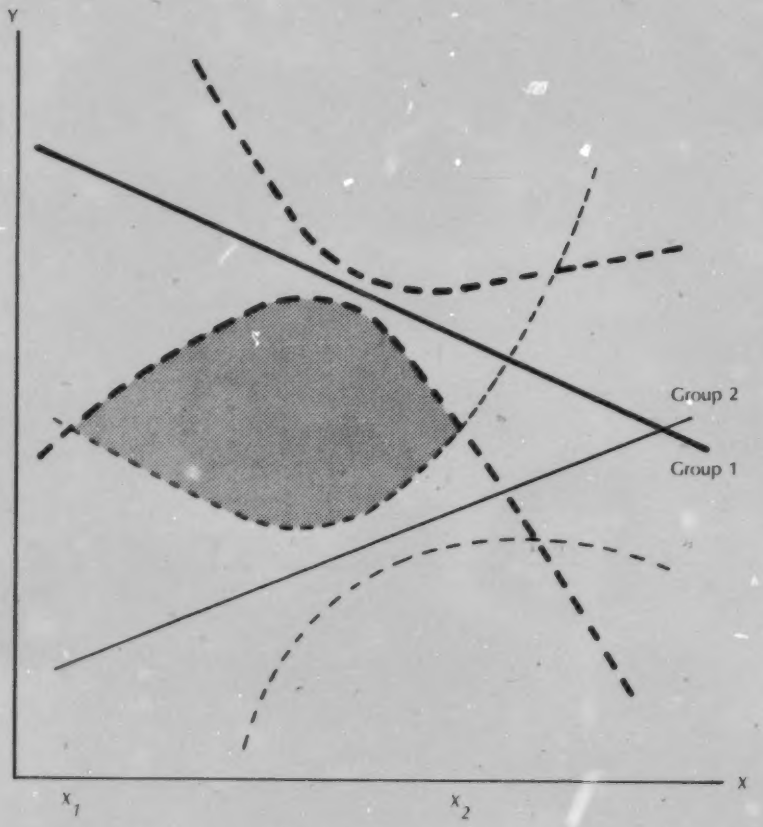


Figure 6c. *Confidence limits in the single-region case. The heavier lines represent Group 1, while the lighter lines represent Group 2. Solid lines are regression lines, while the broken lines are confidence limits. The confidence intervals around the two regression lines overlap except in the shaded area; thus X_1 to X_2 constitutes a single region of significance.*

terest here is the fact that *all* analyses which involved group slope differences significant at or beyond the .05 level produced two regions of significance, while *all* analyses which involved group slope differences significant between .10 and .05 produced either a single region or no region. Thus, a single region is obtained only when the difference between group regressions is marginally significant. These results also allow the speculation that a single region is obtained only when the chance probability of the difference between group regression slopes is numerically greater than the significance level chosen for the regions of significance analysis. In other words, it is possible that (1) setting the same significance level for both the homogeneity of group regressions test and the regions of significance test and (2) applying the latter test only when the former is significant (i.e., when there is acceptable evidence for the existence of an aptitude-treatment interaction) will eliminate the single-region case.

Importance of a Region of Significance in the Single-Aptitude Case

The existence of a region of significance does not necessarily indicate the practical importance of that region. For example, if a region of significance contains no observed data points, then that region is of little importance. (Program ATILIN1 only reports regions of significance which fall within the range of the observed data.) Furthermore, regions of significance are established on the basis of general relationships observed across the entire range of aptitude values. The Johnson-Neyman technique defines a region of significance in terms of differences between group regressions (predicted values) and not on the basis of the observed data within that region. The actual pattern of observed results within a region of significance may be in conflict with the general predicted relationships, and in this case the region would be of little importance. Figure 7 presents a simplified example of this latter situation. Note in Figure 7 that the left region of significance (below point A) is evidenced because the Treatment *O* regression line (predicted scores) is significantly above the Treatment

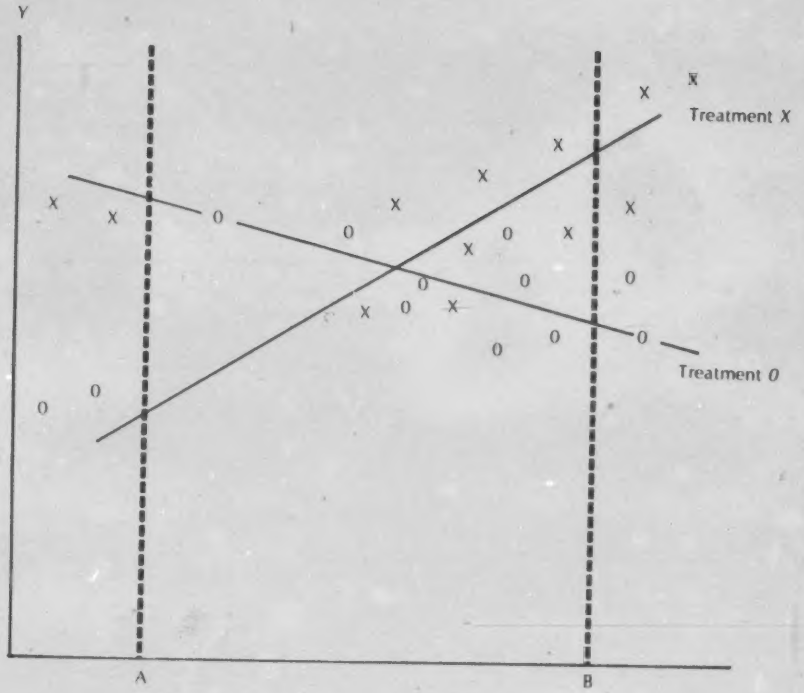


Figure 7. *A region of significance is defined to the left of point A and to the right of point B.*

X regression line (predicted scores). The left region of significance would usually be taken as a region where Treatment *O* was superior. Note, however, that the observed data within this region indicate the exact *opposite* relationship. Any conclusion about the superiority of Treatment *O* within that region is questionable.

A general impression of the importance of a region of significance can be obtained by inspection of the plotted output from Program ATILIN1. If only a small amount of data is contained within a region of significance, then this will be evidenced in the plot. If the pattern of observed results within a region of significance doesn't reflect the general pattern of results (i.e., the overall linear regression lines), then this also should be observable in the plot.

While plotted output provides general impressions about the importance of a region of significance, more objective measures of importance are often desirable. Two measures of the importance of a region of significance can be calculated—(1) the proportion of total observations within a region of significance and (2) an index of the overlap within a region. While ATILIN1 does not provide calculations of these measures, they can be easily determined from the plotted output.

Proportion of total observations within a region of significance. This index of importance is simply the number of observations falling within a region divided by the total number of observations. The greater this proportion, the greater the importance of the region of significance.

Index of overlap within a region of significance. Given a region of significance, we cannot be certain that any given *S* in the group predicted to be superior actually performed better than all the *Ss* in the other group. Some Group 1 *Ss* will perform better than Group 2 *Ss* even though the interaction and region of significance indicate that Group 2's treatment was superior to Group 1's treatment in that region. Figure 8 illustrates such overlap in a region of significance.

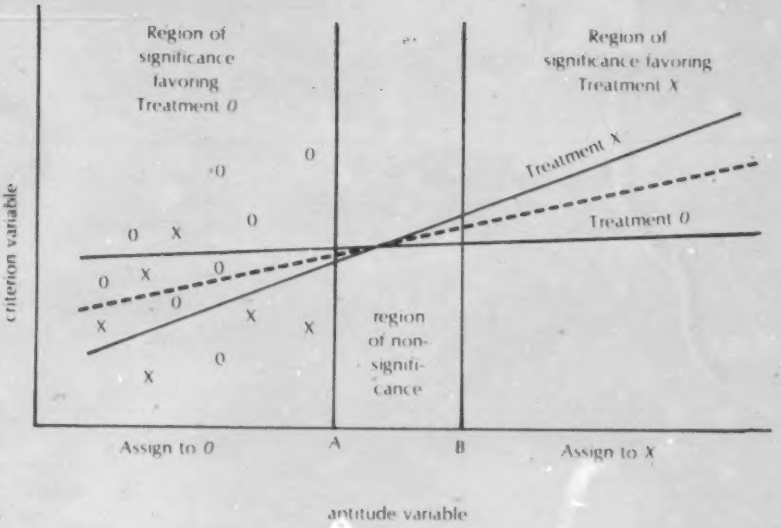


Figure 8. *Overlap within a region of significance.*

Consider the region of significance bounded by point *A* in Figure 8. Notice that, even though Treatment *O* is superior to Treatment *X* for the area which lies to the left of point A, some *X* Ss fall closer to the regression line for Treatment *O* than to the *X* regression line, and that some *O* Ss fall closer to the regression line for Treatment *X* than to the *O* regression line. We can expect such overlapping to occur even when regions of significance are defined at a high level of confidence.

An index of the extent of such overlapping is the percent of all subjects falling within a region of significance who actually demonstrate a criterion score inconsistent with their treatment group. The smaller the value of this index, the greater the importance of the region of significance. Such an index can be calculated by counting the number of subjects in the region of significance who, while assigned to the poorer treatment, actually performed above the midline between the regression lines for the two groups (i.e., a line equidistant from the two group regressions) and adding to this the number of subjects in the region of significance who, while assigned to the better treatment, actually performed below the midline between regressions. The percentage of both types of deviations within a region is calculated by finding the midline between the group regressions and then determining whether each observation falls above or below this line. Let $Mpt(X_i)$ symbolize the midline criterion score for a predictor score of X_i. Note that $[Mpt(X_i), X_i]$ indicates the set of points falling on the midline. The midline between group regressions is given by

$$Mpt(X_i) = \frac{\overline{Y}_1 + b_1(X_i - \overline{X}_1) - \overline{Y}_2 - b_2 X_i - \overline{X}_2)}{2} + \overline{Y}_2 + b_2(X_i - \overline{X}_2)$$

or, simplifying,

$$Mpt(X_i) = \frac{\overline{Y}_1 + b_1(X_i - \overline{X}_1) + \overline{Y}_2 + b_2 X_i - \overline{X}_2)}{2}$$

where $\overline{Y}_1$, $\overline{X}_1$ and b_1 represent the criterion mean score, aptitude mean score and regression coefficient (criterion on aptitude), respectively, for one treatment and $\overline{Y}_2$, $\overline{X}_2$ and b_2, these same values for the other treatment. For subject n with criterion score Y_n and aptitude

score X_n, the distance from the midline is given by

$$D = Y_n - Mpt(X_n).$$

D will be zero when the observation falls on the midline, positive when it falls above it, and negative when it falls below it. D's for observations of the better treatment are expected to be positive and D's for observations of the poorer treatment are expected to be negative. Exceptions are considered "misses" and are tallied and reported as a percent of the total number of observations within the region. In Figure 8, two observations (O's) from the better treatment fell below the midline and two observations (X's) from the poorer treatment fell above it. Both types of "misses" constitute 28 percent of the observations that lay within the region of significance. We, therefore, would report a 28-percent overlap for the region of significance bounded by the aptitude value A. A small amount of overlap indicates that the relationships among the data actually observed within the region are consistent with the predicted relationships used to establish the existence of that region of significance. A large amount of overlap indicates that the observed data contradict the validity of a region of significance. The greater the overlap, the less the importance of the region of significance.

It is important to note that a subject from Treatment 1 scoring closer to the regression line for Treatment 2 does not provide information as to whether that subject has been assigned to a treatment incorrectly. This becomes obvious when we consider a subject who is assigned to the better treatment within a region of significance but whose score falls, let us say, at or below the regression for the poorer treatment in this region. Such a S may be already performing the best that can be expected from either of the treatments and placing him in the opposing treatment might depress his criterion score below even its present level. The investigator cannot infer that the assignment of overlapping subjects to any other treatment would necessarily bring the data into better fit with the overall regression lines.

Confidence intervals for the differences between group regressions. Cronbach and Snow (1974) have developed an alternative procedure relevant to the importance of aptitude-treatment interaction results. Cronbach and Snow suggest the calculation of confidence intervals for the difference between regression lines at all values of the predictor variable. As Figure 9 illustrates, Cronbach and Snow's confidence region will be narrowest at the mean of the aptitude variable and widen to either side. Cronbach and Snow's technique is essentially a confidence interval technique for the differences between means. A direct statement about the limits of the interaction effect is attained by setting confidence limits on the population differences corresponding to the differences in outcome that describe a sample interaction. Such confidence intervals put the differences between regression slopes in proper perspective in that they demonstrate the probable range of real differences as is shown by the hyperbola in Figure 9.

The equation for the confidence limits hyperbola is given by

$$\delta = t_{2,df}\left[\frac{1}{N_A} + \frac{1}{N_B} + \frac{(X-\bar{X}_{(A)})^2}{N_A - 1}\left(\frac{1}{s^2_{X(A)}}\right) + \frac{(X-\bar{X}_{(B)})^2}{N_B - 1}\left(\frac{1}{s^2_{X(B)}}\right)\right] s^2_e \Bigg)^{1/2}$$

where N_A is the number of subjects in Treatment A; N_B, the number of subjects in Treatment B; X, the aptitude variable value; $\bar{X}_{(A)}$ and $\bar{X}_{(B)}$, the aptitude means for the two treatments; $s^2_{X(A)}$ and $s^2_{X(B)}$, the aptitude variances for the two treatments; $t_{2,df}$, the tabled t-value for 2 and df degrees of freedom and the $1 - \alpha$ confidence level; s^2_e, the criterion residual mean square (mean square for the criterion deviations from the treatment regression lines, pooled over treatments); and df, the degrees of freedom for s^2_e. $\Delta\hat{Y} \pm \delta$, a function of X (the aptitude variable), gives the hyperbola that describes the confidence limits. The values of $\Delta\hat{Y}$ describe the observed interaction and these values are obtained by subtracting one within-treatment

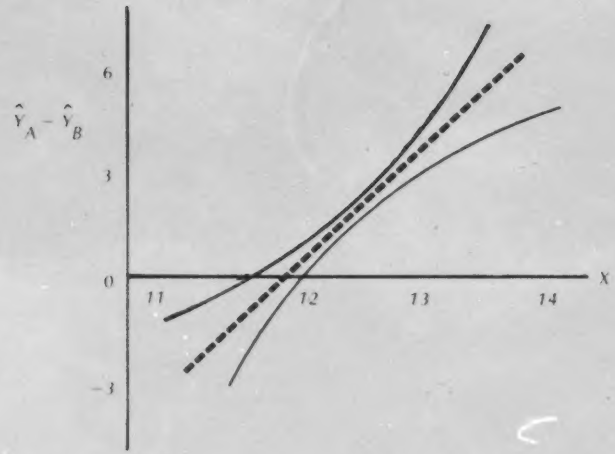


Figure 9. *A simultaneous confidence interval around the difference* $(\hat{Y}_A - \hat{Y}_B)$ *in group regressions. Regressions intersect, i.e.,* $\hat{Y}_A - \hat{Y}_B = 0$, *at* $X = 12$.

regression equation from the other (i.e., $\Delta\hat{Y} = \hat{Y}_A - \hat{Y}_B$).

Cronbach and Snow refer to this calculation as a simultaneous confidence limit in that it is defining a confidence interval for all values of *X*. This approach is somewhat more conservative than the successive confidence interval noted by Potthoff (1964), as the latter will lead to a larger confidence interval and will fall out further toward both extremes of the distribution than will Cronbach and Snow's procedure.

References

Borich, G. D., Godbout, R. C., Peck, R. F., Kash, M. M., & Poynor, L. H. *Final report: An evaluation of the personalized model of teacher training.* Austin, Texas: Research and Development Center for Teacher Education, The University of Texas at Austin, 1974.

Cronbach, L. J., & Snow, R. E. *Aptitudes and instructional methods.* Stanford: Stanford University, College of Education, 1974 (mimeo).

Johnson, P. O., & Jackson, R. W. B. *Modern statistical methods.* Chicago: Rand McNally, 1959.

Johnson, P. O., & Neyman, J. Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs*, 1936, *1*, 57-93.

Kerlinger, F. N., & Pedhazur, E. J. *Multiple regression in behavioral research.* New York: Holt, Rinehart & Winston, 1973.

Potthoff, R. F. On the Johnson-Neyman technique and some extensions thereof. *Psychometrika*, 1964, *29*, 241-255.

Walker, H., & Lev, J. *Statistical inference.* New York: Holt, Rinehart & Winston, 1953, 387-412.

CHAPTER IV

ATILIN2

Program Description

This program tests homogeneity of group regressions and defines regions of significance for the case in which there are two treatment groups and two continuous aptitudes or traits that are linearly related to a criterion. Program outputs (1) table of summary statistics—group sizes, means, standard deviations, correlations between aptitudes and criterion and the intercorrelation between aptitudes; (2) multiple regression equations (*Y*-intercepts and regression coefficients) for each group; (3) points at which the line of nonsignificance intersects the *X* (first aptitude) and *Z* (second aptitude) axes and its slope; (4) *F*-value, degrees of freedom and probability for the homogeneity of group regressions test for (a) both aptitudes simultaneously and (b) each aptitude separately; (5) *F*-value, degrees of freedom and probability for the test of common intercepts (analysis of covariance); and (6) equation for the region(s) of significance. A flow chart for program ATILIN2 is presented in Figure 10.

Program Input

Card 1 alphanumeric title card Col 1-80

Card 1 parameter card

Col 1-5	*N* for Group 1 (maximum = 200)
Col 6-10	*N* for Group 2 (maximum = 200)
Col 12	missing data option 0 = all data valid 1 = blanks are invalid 2 = blanks and zeroes are invalid
Col 14	output option 0 = plot 1 = film 2 = printed output only

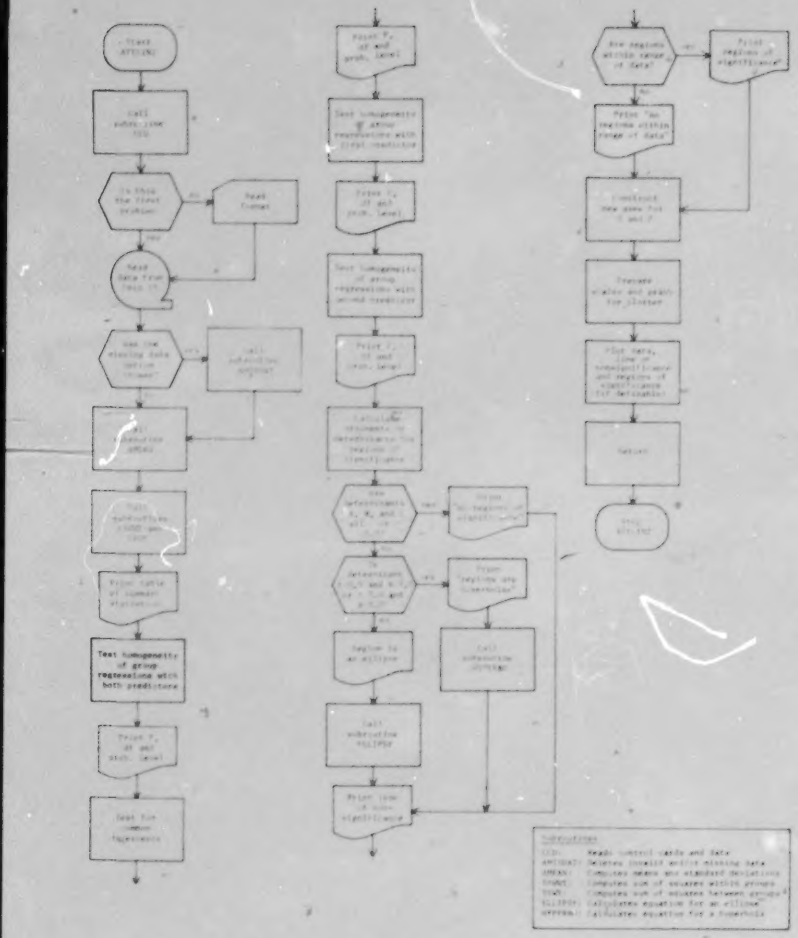


Figure 10. *Program ATILIN2 flow chart.*

	Col 16	option for table of predictor and criterion scores listed by subject within treatments. If this option is taken, ID codes will be read according to format and will be printed out (A5) along with corresponding predictor and criterion scores for subjects in each treatment group. Subjects with missing data will not be listed in this table. 0 = no table 1 = list ID codes and scores (begin format cards with A mode field)
	Col 18	no. of cards per subject in Group 1
	Col 20	no. of cards per subject in Group 2
	Col 25-34	alpha level for regions of significance
Card 3	format for Group 1 Col 1-80 followed by Group 1 data	
Card 4	format for Group 2 Col 1-80 followed by Group 2 data	
Card 5	blank (after last problem) for multiple problems repeat cards 1-4, omitting data	

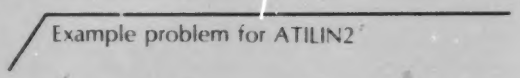
Data cards should contain subject ID codes (if desired), the two aptitude scores, and then the criterion score. If Col 16 on the parameter card is 0, then formats must specify three F-mode fields—the first two for the two aptitudes and the third for the criterion. If Col 16 on the parameter card is 1, then formats must specify an initial A-mode field (A5 or less) for the ID code and then the three F-mode fields.

Example Problem

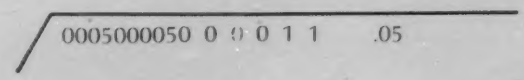
Data for this problem will be the first and second predictors (aptitudes) and the criterion given as sample data in Chapter VII (p. 95) of this manual. Program control cards for this example problem are as follows.

52

1. Alphanumeric title card

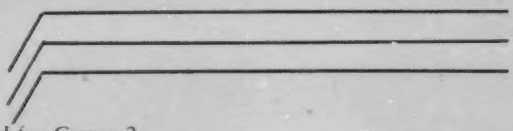


2. Parameter card

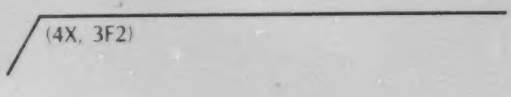


3. Format card for Group 1

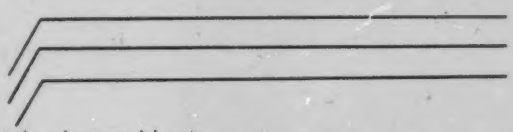
4. Group 1 data



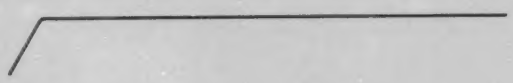
5. Format card for Group 2



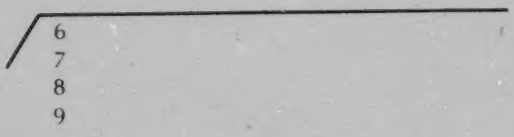
6. Group 2 data



7. Blank card (after last problem)



8. EOF



Printed output for this example problem is given in Figure 11 and plotted output is given in Figure 12.

ATILIN2 SUMMARY STATISTICS

STATISTICS	GROUP 1	GROUP 2
N	50	50
MEAN		
CRITERION	50.0000	48.5000
PREDICTOR 1	54.1000	54.8000
PREDICTOR 2	50.6000	48.5000
SIGMA		
CRITERION	20.1742	19.9812
PREDICTOR 1	20.6807	26.1526
PREDICTOR 2	20.5339	19.9812
CORRELATION		
PREDICTOR 1/CRITERION	.9048	-.7900
PREDICTOR 2/CRITERION	-.9439	.6080
PREDICTOR 1 WITH 2	-.8759	-.6885

THE REGRESSION EQUATION FOR GROUP 1 IS Y = 64.6289 + .3271 X + -.6387 Z

THE REGRESSION EQUATION FOR GROUP 2 IS Y = 72.1484 + -.5394 X + .1219 Z

THE F-VALUE FOR THE TEST OF HOMOGENEITY OF GROUP REGRESSIONS IS 139.8168 WITH 2 AND 94 DEGREES OF FREEDOM WHICH HAS A PROBABILITY OF .0000

Figure 11. *Printed output for ATILIN2 example problem—summary statistics and significance tests.*

```
THE F-VALUE FOR THE TEST OF COMMON INTERCEPTS, ASSUMING HOMOGENEITY OF REGRESSION PLANES
IS    .3554 WITH 1 AND   96 DEGREES OF FREEDOM WHICH HAS A PROBABILITY LEVEL OF .5596

CONSIDERING ONLY THE FIRST PREDICTOR, THE F-VALUE FOR THE TEST OF HOMOGENEITY OF GROUP REGRESSIONS
IS   30.4527 WITH 1 AND   94 DEGREES OF FREEDOM WHICH HAS A PROBABILITY OF .0000

CONSIDERING ONLY THE SECOND PREDICTOR, THE F-VALUE FOR THE TEST OF HOMOGENEITY OF GROUP REGRESSIONS
IS   20.1245 WITH 1 AND   94 DEGREES OF FREEDOM WHICH HAS A PROBABILITY OF .0001


THE REGION OF SIGNIFICANCE IS A HYPERBOLA WITH AN   EQUATION
X-SQUARED /   -79.7879 + Z-SQUARED /     1.+.3689 = 1.


THE LINE OF NON-SIGNIFICANCE INTERSECTS THE X-AXIS AT    8.6866 AND THE Z-AXIS
AT   -9.8965 AND HAS A SLOPE OF     1.+.1393


END OF JOB.
```

Figure 11 (continued). *Printed output for ATTLIN2 example problem—summary statistics and significance tests.*

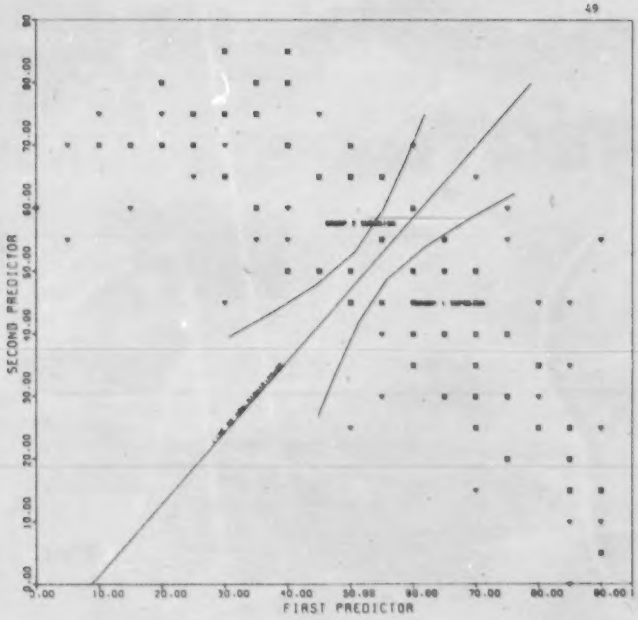
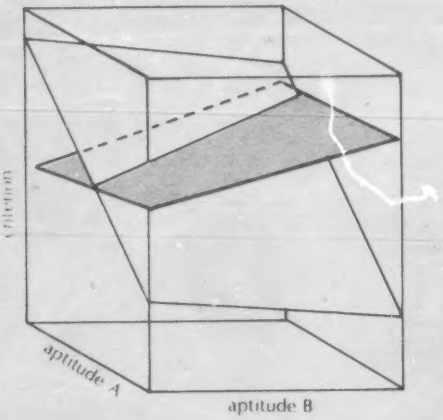


Figure 12. *Plotted output for ATILIN2 example problem.*

ATILIN2

1. Homogeneity of Group Regressions, Multiple Aptitudes

By extending the linear single-aptitude model used in ATILIN1, ATILIN2 simultaneously tests pairs of aptitudes and isolates specific aptitudes for which there are unequal slopes. Borich (1972) and Johnson and Jackson (1959) provide discussions of aptitude-treatment interactions involving pairs of aptitudes. For more than one aptitude, regression planes and hyperplanes (three or more aptitudes) are analogous to the regression lines of the single-aptitude case. For two groups and two aptitudes, the linear model may be extended to fit the following case in which two aptitudes are linearly related to a criterion.



ATILIN2 constructs a full model of the form:

$$Y_i = b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + b_4X_{4i} + b_5Z_{1i} + b_6Z_{2i} + e_i \quad [3]$$

where b_1 is the regression coefficient for the first group membership vector, X_{1i} (scored 1 if S_i is in Group 1, scored 0 if not); b_2, the regression coefficient for the second group membership vector, X_{2i} (scored

1 if S_i is in Group 2, scored 0 if not); b_3, the regression coefficient of the product (X_{3i}) of X_{1i} and the first aptitude vector; b_4, the regression coefficient of the product (X_{4i}) of X_{2i} and the first aptitude vector; b_5 and b_6, regression coefficients of the products (Z_{1i}, Z_{2i}) of X_{1i} and X_{2i}, respectively, and the second aptitude vector. The residual sum of squares for this full model (Σe_i^2) has $N - 6$ degrees of freedom or, for more than two treatment groups, $N - 3k$ degrees of freedom where k is the number of treatment groups.

Program ATILIN2 performs three significance tests relevant to homogeneity of group regressions. These three significance tests involve comparison of the full model [3] to three different restricted models. Each of these significance tests is discussed in turn.

Test 1—Simultaneous test of slope differences on both aptitudes. To test if there are parallel regression plane slopes (criterion regressed simultaneously on both aptitudes) for the two treatments, a restricted model is formed by setting b_3 equal to b_4 and setting b_5 equal to b_6 in the full model [3]. The resulting model is:

$$Y_i = b_1X_{1i} + b_2X_{2i} + b_3(X_{3i} + X_{4i}) + b_5(Z_{1i} + Z_{2i}) + f_i \quad [4]$$

or, equivalently,

$$Y_i = b_1X_{1i} + b_2X_{2i} + b_3A_{1i} + b_5A_{2i} + f_i$$

where A_{1i} is the vector of scores for the first aptitude and A_{2i} is the vector of scores for the second aptitude. The residual sum of squares for this restricted model (Σf_i^2) has $N - 4$ *df* or, for more than two treatment groups, $N - k - 2$ *df* where k equals the number of treatment groups. The *F*-test for differences in group regression plane slopes is constructed in the usual manner using the error sums of squares from the full [3] and restricted [4] models.

Test 2—Test of slope differences on Aptitude 1 with slope differences on Aptitude 2 covaried. To test if there are regression slope differences with regard to Aptitude 1 over and above slope differences attributable to Aptitude 2, a restricted model is formed by

setting b_3 equal to b_4 in the full model [3]. The resulting restricted model is:

$$Y_i = b_1X_{1i} + b_2X_{2i} + b_3(X_{3i} + X_{4i}) + b_5Z_{1i} + b_6Z_{2i} + g_i \quad [5]$$

or, equivalently,

$$Y_i = b_1X_{1i} + b_2X_{2i} + b_3A_{1i} + b_5Z_{1i} + b_6Z_{2i} + g_i.$$

The residual sum of squares for this restricted model (Σg_i^2) has $N - 5$ *df* or, for more than two treatment groups, $N - 2k - 1$ *df* where k is the number of treatment groups. The *F*-test for Aptitude 1 slope differences with Aptitude 2 slope differences covaried is constructed in the usual manner using the error sums of squares from the full [3] and restricted [5] models.

Test 3—Test of slope differences on Aptitude 2 with slope differences on Aptitude 1 covaried. The restriction for this test is $b_5 = b_6$ and the restricted model is:

$$Y_i = b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + b_4X_{4i} + b_5(Z_{1i} + Z_{2i}) + h_i \quad [6]$$

or, equivalently,

$$Y_i = b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + b_4X_{4i} + b_5A_{2i} + h_i.$$

The error sum of squares (Σh_i^2) has $N - 5$ *df* for two treatment groups or $N - 2k - 1$ *df* for k treatment groups. The *F*-test for Aptitude 2 slope differences over and above slope differences attributed to Aptitude 1 is constructed with [3] as the full model and [6] as the restricted model.

Interpretation of results—Aptitudes uncorrelated. When the two aptitudes are not significantly correlated, the results of the three significance tests are easily interpreted. Given significant results from Test 1 (the simultaneous test), an aptitude-treatment interaction exists. Tests 2 and 3 can then be examined to determine if the interaction involves Aptitude 1, Aptitude 2, or both. If Test 2 is significant,

then Aptitude 1 is involved in the interaction and consideration of Aptitude 1 is necessary to an adequate description of the obtained interaction. If Test 2 is nonsignificant, then Aptitude 1 is not involved in the interaction and need not be considered in describing the interaction. Parallel conclusions regarding Aptitude 2 can be arrived at on the basis of Test 3.

Interpretation of results—Aptitudes correlated. Consider the case in which the two aptitudes are significantly correlated. Significant results from Test 1 again indicate the existence of an aptitude-treatment interaction. However, the confounding of the two aptitudes complicates interpretation of the results of Test 2 and Test 3. For correlated aptitudes, conclusions must be based upon simultaneous consideration of the results of Test 2 and Test 3. If both tests yield significance, then both aptitudes must be considered in order to adequately describe the obtained interaction. If Test 2 is significant and Test 3, nonsignificant, then Aptitude 1, considered by itself, allows an adequate description of the interaction. In this latter case, consideration of Aptitude 2 would be redundant. If Test 2 is nonsignificant and Test 3, significant, then it is sufficient to consider a single aptitude, Aptitude 2.

When aptitudes are correlated, it is quite possible for an interaction to exist (significant Test 1 results) but for both Test 2 and Test 3 to yield nonsignificant results. Such a case arises when the aptitude variance involved in the interaction is variance held in common by the two confounded aptitudes and neither Aptitude 1 nor Aptitude 2 is uniquely involved in the interaction. When neither Test 2 nor Test 3 provides significance, then it is sufficient to consider a single aptitude, but the choice of which aptitude to consider is arbitrary. When aptitudes are highly correlated there is little need to include both aptitudes in the analysis as the variance which is explained by one aptitude is also explained by the other. Therefore, aptitude pairs for which the above sequence of tests is most applicable are those in which the aptitudes are minimally related to each other and each is significantly related to the criterion.

2. *Test of Common Intercepts (Analysis of Covariance)*

To test the null hypothesis that two treatment groups are not significantly different when subjects score at the mean of both aptitudes, ATILIN2 constructs a full model of the form

$$Y_i = b_1X_{1i} + b_2X_{2i} + b_3A_{1i} + b_4A_{2i} + e_i \qquad [7]$$

where Y_i is the criterion; b_1, the regression coefficient for the first group membership vector (X_{1i}); b_2, the regression coefficient for the second group membership vector (X_{2i}); b_3, the regression coefficient for the first aptitude (A_{1i}); and b_4, the regression coefficient for the second aptitude (A_{2i}). The residual sum of squares for this full model (Σe_i^2) has $N - 4$ *df* or, for more than two treatment groups, $N - k - 2$ *df* where k equals the number of treatment groups.

The test of common intercepts involves placing the restriction, $b_1 = b_2$, on the full model. The resulting restricted model is

$$Y_i = a + b_3A_{1i} + b_4A_{2i} + f_i \qquad [8]$$

where a is the regression constant, f_i is the error vector, and the other terms are defined the same as in the case of the full model. The error sum of squares for the restricted model (Σf_i^2) has $N - 3$ *df*. The *F*-test for common intercepts is constructed in the usual manner using the error sums of squares from full [7] and restricted [8] models.

3. *Line of Nonsignificance*

The line of nonsignificance is the line of zero difference between the two group regression planes (criterion on both aptitudes)—i.e., it is the intersection of the two regression planes. The regression plane for Treatment 1 is given by

$$Y = a_1 + b_{11}A_1 + b_{21}A_2 + e_1 \qquad [9]$$

and that for Treatment 2 is given by

$$Y = a_2 + b_{12}A_1 + b_{22}A_2 + e_2 \cdot \qquad [10]$$

In [9] and [10], Y represents the criterion; a_1 and a_2, the regression constants for the two treatments; b_{11} and b_{12}, the Aptitude 1 (A_1) regression coefficients for the two treatments; b_{21} and b_{22}, the Aptitude 2 (A_2) regression coefficients for the two treatments. The values of the regression parameters in [9] and [10] can be estimated from the data and these estimates can then be used to write the equation for the line of nonsignificance. The equation for the line of nonsignificance is given by

$$(a_1 - a_2) + (b_{11} - b_{12})A_1 + (b_{21} - b_{22})A_2 = 0.$$

4. *Regions of Significance*

When an aptitude-treatment interaction exists, it is of special importance to determine if there are significant group differences within the range of observed aptitude values. A region of significance consists of a set of aptitude values for which predicted criterion performance is significantly different for the two treatments. In the two-aptitude case, regions of significance are defined in the two-dimensional space created by Aptitude 1 and Aptitude 2.

Consider a single point in the two-dimensional space for Aptitude 1 and Aptitude 2. The predicted criterion difference (D) between groups at that point is significant if

$$D^2 \geq F\alpha(P + Q)S_e^2 / (N_1 + N_2 - 6)$$

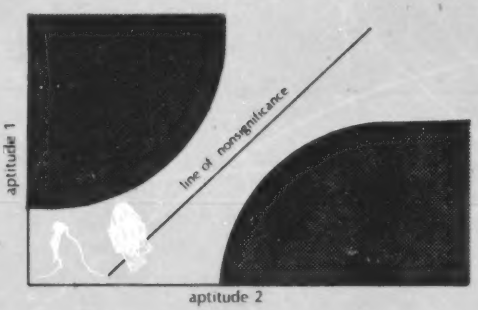
where $F\alpha$ is the F-ratio ($df = 1$ and $N_1 + N_2 - 6$) required for the α level of significance; $(P + Q)$ is a function of (a) the values of Aptitude 1 and Aptitude 2 defining the point in question, (b) the number of subjects in Treatment 1 (N_1), (c) the number of subjects in Treatment 2 (N_2), (d) the means and variances of the two aptitude

variables, (e) the correlation between the two aptitude variables and (f) the correlations of the criterion with each of the aptitude variables; and S_e^2 is the error sum of squares for model [3]. The actual expression for $(P + Q)$ can be found in Johnson and Jackson (1959, p. 443).

Bounding values for regions of significance are given by the following expression:

$$D^2 - \frac{F\alpha(P + Q)S_e^2}{N_1 + N_2 - 6} = 0. \qquad [11]$$

Expression [11] is an equation of the second degree involving A_1^2, A_2^2, A_1A_2, A_1, and A_2 terms where A_1 is Aptitude 1 and A_2 is Aptitude 2. As in the single-aptitude case, one, two or no regions of significance may occur. When there are two regions of significance, [11] defines an hyperbola. How regions of significance fall with regard to this hyperbola can be seen in the following diagram.



The shaded portions of this diagram represent the regions of significance. When there is a single region of significance, [11] defines an ellipse with the region of significance falling within that ellipse.

In a previous discussion of regions of significance defined with regard to a single aptitude (pp.28-33), it was concluded that the single-region finding may be of little importance. In that discussion, it was pointed out that a single region occurred only when there was marginal evidence (homogeneity of group regressions test) for interaction and perhaps only when the chance probability for the

homogeneity of regressions test exceeded the α-level chosen to establish regions of significance. It is suspected that the conditions requisite to finding a single region with two aptitudes are analogous. Thus, the single region finding with regard to two aptitudes may be of little importance.

Importance of the region of significance in the two-aptitude case. Regions of significance mathematically defined in a two-aptitude space can have little or no practical importance. The importance of a region of significance in the two-aptitude case is (1) a positive function of the proportion of total observations which fall within that region and (2) a negative function of the amount of overlapping between the treatments within that region. Both of these indices of importance have been discussed with regard to the single-aptitude case (pp. 33-38). Generalization of the first index (proportion of total observations within the region) and the second index (overlap index) to the two-aptitude case is relatively easy. With two aptitudes, a Treatment 1 observation evidences overlap if that observation falls closer to the Treatment 2 regression plane than the Treatment 1 regression plane and vice versa for a Treatment 2 observation. In other words, an observation is counted as overlapping if it lies on the "wrong" side of the midplane* between the group regression planes. The overlap index is then the number of overlapping observations in a region divided by the total number of observations in that region.

*The midplane equation is

$$Mpt(A_{1i}, A_{2i}) =$$
$$[\bar{Y}_1 + b_{11}(A_{1i} - \bar{A}_{11}) + b_{21}(A_{2i} - \bar{A}_{21}) +$$
$$\bar{Y}_2 + b_{12}(A_{1i} - \bar{A}_{12}) + b_{22}(A_{2i} - \bar{A}_{22})]/2$$

where $Mpt(A_{1i}, A_{2i})$ is the midplane criterion score for Aptitude 1 equal to A_{1i} and Aptitude 2 equal to A_{2i}; $\bar{Y}_1$ and $\bar{Y}_2$ are the criterion means for the two treatments; the b's are from [9] and [10]; $\bar{A}_{11}$ and $\bar{A}_{21}$ are the Treatment 1 means on Aptitude 1 and Aptitude 2; and $\bar{A}_{12}$ and $\bar{A}_{22}$ are the Treatment 2 means on Aptitudes 1 and 2.

References

Borich, G. D. Homogeneity of slopes test for multiple regression equations with reference to aptitude-treatment interactions. *Journal of Experimental Education*, 1972, *40*, 39-42.

Johnson, P.O., & Jackson, R. W. *Modern statistical methods: Descriptive and inductive*. Chicago: Rand McNally, 1959.

CHAPTER V

ATICURV

Program Description

This program tests homogeneity of group regressions and defines regions of significance for the case in which there are two treatment groups and one continuous aptitude or trait that is curvilinearly related to a criterion. Program outputs (1) table of summary statistics—group sizes, means, standard deviations and correlations (Pearson product-moment) between aptitude and criterion; (2) regression equations (Y-intercepts and regression coefficients) for each group; (3) probability that the relationship between the aptitude and criterion is curvilinear for each group; (4) *F*-value, degrees of freedom and probability for the homogeneity of curvilinear regressions test; (5) *F*-value, degrees of freedom and probability for the test of common intercepts for curvilinear data (analysis of covariance); (6) aptitude value(s) at which the curvilinear regressions intersect; and (7) aptitude values which define the region(s) in which treatment groups are significantly different (regions of significance). Figure 13 presents the flow chart for program ATICURV.

Program Input

Card 1 alphanumeric title card Col 1-80

Card 2 parameter card

Col 1-5	*N* for Group 1 (maximum = 200)
Col 6-10	*N* for Group 2 (maximum = 200)
Col 12	missing data option 0 = all data valid 1 = blanks are invalid 2 = blanks and zeroes are invalid
Col 14	output option 0 = plot 1 = film 2 = printed output only

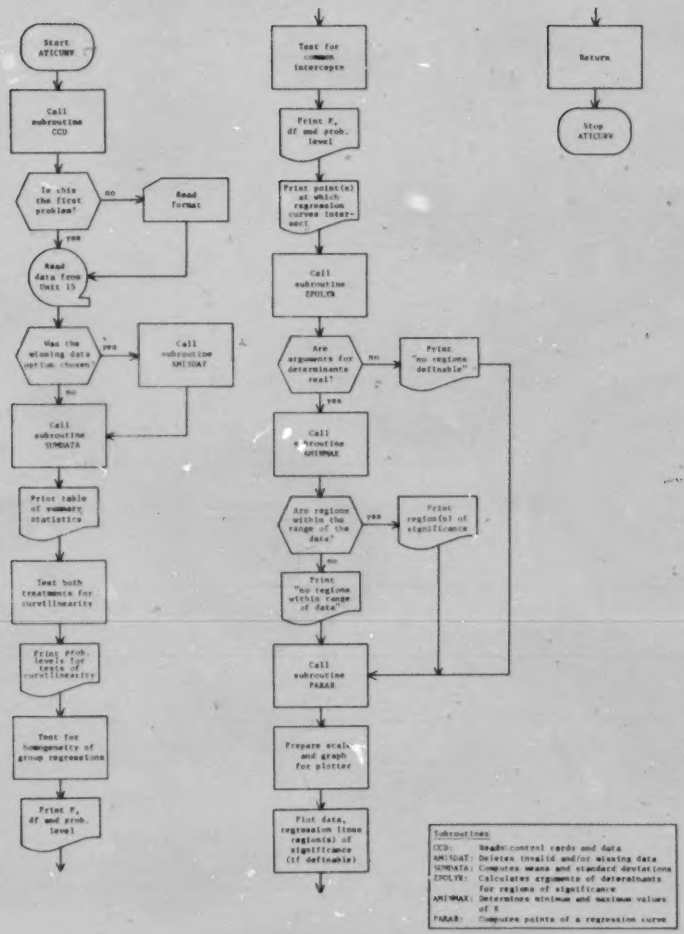

Start ATICURV
Call subroutine CCD
Is this the first problem?
no
Read format
yes
Read data from Unit 15
Was the missing data option chosen
yes
Call subroutine AMISDAT
no
Call subroutine SUMDATA
Print table of summary statistics
Test both treatments for curvilinearity
Print prob. levels for tests of curvilinearity
Test for homogeneity of group regressions
Print F, df and prob. level
Test for common intercepts
Print F, df and prob. level
Print point(s) at which regression curves intersect
Call subroutine ZPOLYR
Are arguments for determinants real?
no
Print "no regions definable"
yes
Call subroutine AMINMAX
Are regions within the range of the data?
yes
Print region(s) of significance
no
Print "no regions within range of data"
Call subroutine PARAB
Prepare scale and graph for plotter
Plot data, regression lines region(s) of significance (if definable)
Return
Stop ATICURV
Subroutines
CCD: Reads control cards and data
AMISDAT: Deletes invalid and/or missing data
SUMDATA: Computes means and standard deviations
ZPOLYR: Calculates arguments of determinants for regions of significance
AMINMAX: Determines minimum and maximum values of x
PARAB: Computes points of a regression curve

Figure 13. *Flow chart for program ATICURV.*

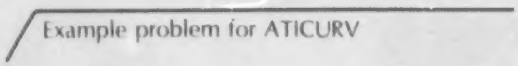
	Col 16	option for table of predictor and criterion scores listed by subject within treatments. If this option is taken, ID codes will be read according to format and will be printed out (A5) along with corresponding predictor and criterion scores for subjects in each treatment group. Subjects with missing data will not be listed in this table. 0 = no table 1 = list ID codes and scores (begin format with A mode field)
	Col 18	no. of cards per subject in Group 1
	Col 20	no. of cards per subject in Group 2
	Col 25-34	alpha level used to test whether a linear or curvilinear model is appropriate
	Col 35-44	alpha level used to determine regions of significance
Card 3	format for Group 1 Col 1-80 followed by Group 1 data	
Card 4	format for Group 2 Col 1-80 followed by Group 2 data	
Card 5	blank (after last problem) for multiple problems repeat cards 1-4, omitting data	

Data cards should contain subject ID codes (if desired), the aptitude score, and then the criterion score. If Col 16 on the parameter card is 0, then formats must specify two F-mode fields—the first field for the aptitude and the second for the criterion. If Col 16 on the parameter card is 1, then formats must specify an initial A-mode field (A5 or less) for the ID code and then the two F-mode fields.

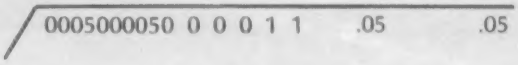
Example Problem

Data for this problem will be the first predictor and the criterion given as sample data in Chapter VII (p. 95) of this manual. Program control cards for this example problem are as follows:

1. Alphanumeric title card

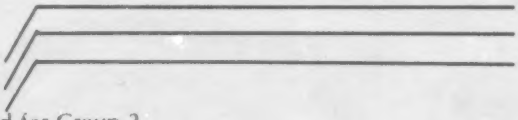


2. Parameter card



3. Format card for Group 1

4. Group 1 data

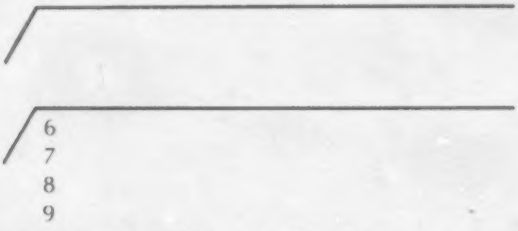


5. Format card for Group 2

6. Group 2 data

7. Blank card (after last problem)

8. EOF



Printed output for this example problem is given in Figure 14 and plotted output is given in Figure 15.

```
        ATICURV SUMMARY STATISTICS

STATISTICS          GROUP 1          GROUP 2

N                        50               50

MEAN
  CRITERION          50.0000          48.5000
  PREDICTOR          54.1000          54.0000

SIGMA
  CRITERION          20.1742          19.9812
  PREDICTOR          20.6007          20.1526

CORRELATION           .9048            -.7900


A COMPARISON OF A LINEAR AND QUADRATIC MODEL FOR EACH TREATMENT GROUP SUGGESTS CURVILINEARITY
AT THE .920 LEVEL FOR TREATMENT 1 AND AT THE .003 LEVEL FOR TREATMENT 2.


THE REGRESSION EQUATION FOR GROUP 1 IS  Y =   1.6900  +     .9081  X  +    -.0082  XX

THE REGRESSION EQUATION FOR GROUP 2 IS  Y =  67.7431  +     .1939  X  +    -.0081  XX


THE F-VALUE FOR THE TEST OF HOMOGENEITY OF GROUP REGRESSIONS IS  143.9273 WITH 2 AND   94
DEGREES OF FREEDOM WHICH HAS A PROBABILITY OF .0000


THE F-VALUE FOR THE TEST OF COMMON INTERCEPTS, ASSUMING HOMOGENEITY OF GROUP REGRESSIONS
IS     .0664 WITH 1 AND    96 DEGREES OF FREEDOM WHICH HAS A PROBABILITY OF .7938


THE REGRESSION CURVES INTERSECT AT THE POINTS WHERE X IS EQUAL TO   56.885 AND WHERE X
IS EQUAL TO  -147.779


A REGION OF SIGNIFICANCE EXTENDS FROM    0.000  TO    53.224
WHERE    0.0000 IS THE MINIMUM OBSERVED APTITUDE VALUE


A REGION OF SIGNIFICANCE EXTENDS FROM   60.220  TO    90.000
WHERE 90.0000  IS THE MAXIMUM OBSERVED APTITUDE VALUE


END OF JOB.
```

Figure 14. *Printed output for ATICURV example problem—summary statistics and significance tests.*

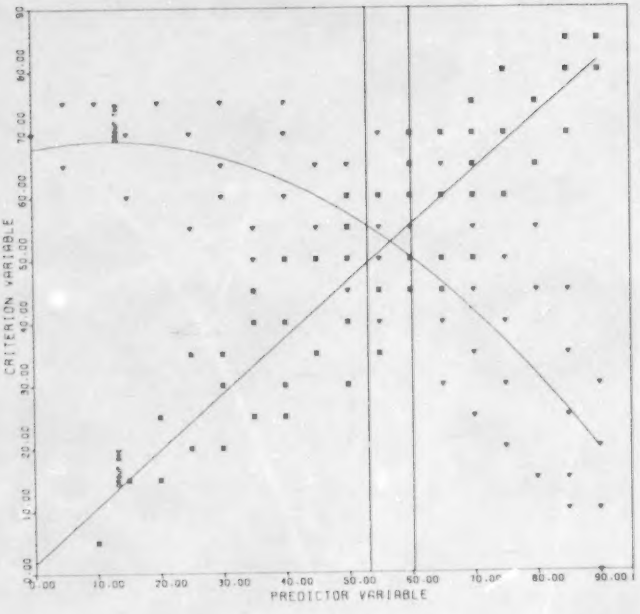


Figure 15. *Plotted output for ATICURV example problem.*

ATICURV

1. Test for Curvilinear Data

A general discussion of the application of regression analysis methods to nonlinear relationships can be found in Kelly, Beggs, McNeil, Eichelberger, and Lyon (1969). To test the hypothesis that the within-group relationship between an aptitude and a criterion is curvilinear, the standard curvilinear (quadratic) prediction model is constructed:

$$Y = a + b_1 X_1 + b_2 X_1^2 + e \qquad [12]$$

where Y is the criterion; a, the regression constant (intercept); b_1, the regression coefficient for the Group 1 aptitude vector, X_1; and b_2, the regression coefficient for the vector (X_1^2) composed of the squares of the Group 1 aptitude scores. The residual sum of squares (Σe^2) for this model has degrees of freedom equal to the number of subjects minus the number of independent parameters—$N_1 - 3$, where N_1 is the number of subjects in Group 1.

The test for curvilinearity involves comparison of the predictive efficiency of model [12], the curvilinear model, with a simple linear model. The appropriate simple linear model is

$$Y = a + b_1 X_1 + f. \qquad [13]$$

The error sum of squares (Σf^2) for model [13] has degrees of freedom equal to $N_1 - 2$. The F-test for curvilinearity is constructed in the usual manner using the error sums of squares from the full model [12] and restricted model [13]. A significant F-ratio indicates a significant curvilinear regression of the criterion on the aptitude.

ATICURV repeats this test of curvilinearity for Treatment 2 and reports the results of the tests for both treatments. Note that the operation of this program is unintluenced by the results of the tests of curvilinearity. Thus, ATICURV proceeds with subsequent significance testing (homogeneity of regressions, analysis of covariance, and regions of significance) as though both within-treatment regressions are curvilinear, regardless of whether the test of curvilinearity is significant for both treatments, only one treatment, or neither of the treatments. Use of ATICURV is not recommended to the researcher who knows beforehand that he is dealing with linear rather than curvilinear data. Instead, program ATILIN1 should be used in this case. Application of ATICURV to *linear* data can result in a possible loss of statistical power due to the loss of degrees of freedom incurred by including curvilinear variables (squared terms) in the regression models.

2. Homogeneity of Curvilinear Regressions

To test the hypothesis that the regressions (including curvilinear components) for the two treatments are parallel, the following full model is constructed:

$$Y_i = b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + b_4X_{4i} + b_5X_{3i}^2 + b_6X_{4i}^2 + e_i \qquad [14]$$

where X_{1i} is the first group membership vector (scored 1 if Subject i is in Group 1 and scored 0 otherwise); X_{2i}, the second group membership vector (scored 1 if Subject i is in Group 2 and scored 0 otherwise); X_{3i}, the product of X_{1i} and the aptitude vector; and X_{4i}, the product of X_{2i} and the aptitude vector. The b's in [14] are regression coefficients and e_i represents the error vector. The error sum of squares for this full model has $N - 6$ *df* or, for more than two treatment groups, $N - 3k$ *df* where N is the total number of subjects and k is the number of treatment groups.

To test if the group regressions are homogeneous (parallel), the

restrictions $b_3 = b_4$ and $b_5 = b_6$ are placed on the full model. The resulting restricted model is

$$Y_i = b_1X_{1i} + b_2X_{2i} + b_3(X_{3i} + X_{4i}) + b_5(X_{3i}^2 + X_{4i}^2) + f_i$$

or, equivalently,

$$Y_i = b_1X_{1i} + b_2X_{2i} + b_3A_i + b_5A_i^2 + f_i \qquad [15]$$

where A_i is the vector of aptitude scores and A_i^2 is the vector of the squares of the aptitude scores. The error sum of squares (Σf_i^2) for this restricted model has $N - 4$ *df* or, for more than two treatment groups, $N - k - 2$ *df* where k is the number of treatment groups.

The *F*-ratio for homogeneity of curvilinear group regressions is constructed in the usual manner using the error sums of squares from the full [14] and restricted [15] models. A significant *F*-ratio indicates that the curvilinear regressions differ for the two treatment groups and thus that an aptitude-treatment interaction exists.

3. Tests of Common Intercepts (Curvilinear Analysis of Covariance)

To test the hypothesis that two treatment groups are significantly different when subjects score at the mean of an aptitude which is curvilinearly related to the criterion, model [15] is employed as the full model. The restriction $b_1 = b_2$ is placed on [15], resulting in the following restricted model:

$$Y_i = a + b_3A_i + b_4A_i^2 + g_i \qquad [16]$$

where a is the regression constant (*Y*-intercept); g_i is the residual error; and the other terms are as defined for [15]. The error sum of squares (Σg_i^2) for this restricted model has $N - 3$ *df*. The *F*-test for significant *Y*-intercept differences for the two treatments is constructed

in the usual manner using the error sums of squares from the full [15] and restricted [16] models. A significant *F*-ratio indicates that the *Y*-intercepts differ for the two treatments. Given that the *Y*-intercept for one treatment is significantly greater than that of the other and given that group regressions are homogeneous, it can be concluded that the treatment with the higher intercept demonstrates overall criterion superiority within the range of the observed data.

4. Points of Intersection

When group regressions are curvilinear, there may be one, two, or no points at which the regressions for the two groups intersect. These different occurrences are shown in Figure 16.

Mathematically, the point(s) of intersection are determined from the within-treatment regression equations. The regression equation for Treatment 1 is

$$\hat{Y}_1 = a_1 + b_{11}A + b_{21}A^2 \qquad [17]$$

where $\hat{Y}_1$ is the predicted criterion score; a_1, the regression constant; b_{11}, the regression coefficient for the aptitude variable (A); and b_{21}, the regression coefficient for the square of the aptitude variable. The analogous regression equation for Treatment 2 is

$$\hat{Y}_2 = a_2 + b_{12}A + b_{22}A^2. \qquad [18]$$

A point of intersection occurs when $\hat{Y}_1 - \hat{Y}_2 = 0$ or, equivalently, when

$$(a_1 - a_2) + (b_{11} - b_{12})A + (b_{21} - b_{22})A^2 = 0. \qquad [19]$$

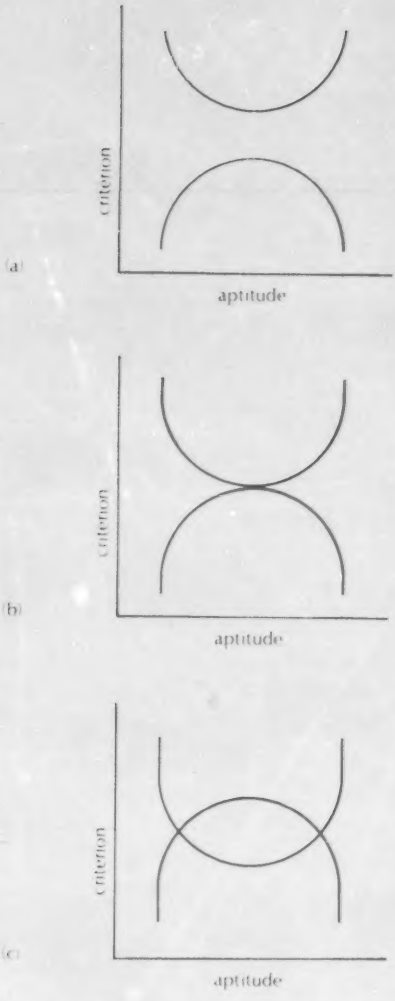


Figure 16. *Points of intersection for three hypothetical cases. Curved lines represent the regressions for two treatment groups. Case (a) involves no point of intersection; case (b), one point of intersection; and case (c), two points of intersection.*

Solving [19] for A yields aptitude values corresponding to the points of intersection. Expression [19] is a quadratic equation and applying the well-known solution for such an equation produces

$$A = \frac{-(b_{11} - b_{12}) \pm \sqrt{(b_{11} - b_{12})^2 - 4(b_{21} - b_{22})(a_1 - a_2)}}{2(b_{21} - b_{22})} \qquad [20]$$

If the term under the radical in [20] is negative, then there are no real solutions to [20] and there are no points of intersection. If the term under the radical is 0, then [20] provides a single solution and the single point of intersection occurs at the corresponding aptitude value. If the term under the radical is positive, then there are two solutions to [20] corresponding to the two aptitude values at which intersections occur.

5. *Regions of Significance*

Johnson and Neyman (1936), while interested in problems involving linear relationships, employed a statistical paradigm which does not restrict the relationship between a criterion and aptitude variable to one of linearity. Wunderlich and Borich (1974) have extended the regions of significance procedure originally suggested by Johnson and Neyman to problems involving a quadratic relationship between criterion and aptitude. Given an aptitude-treatment interaction and curvilinear (quadratic) regressions within treatments, it is of primary interest to determine if the difference between predicted criterion scores—i.e., the distance between the regression curves—is significant for any aptitude values within the range of observed aptitude scores. A region of significance consists of a set of aptitude values for which predicted criterion performance is significantly different for the two treatments.

Consider a single aptitude value. The predicted criterion difference (D) between treatments at that value is significant if

$$D^2 \geq F\alpha(P + Q)S_e^2 \;/\; (N_1 + N_2 - 6) \qquad [21]$$

where $F\alpha$ is the F-ratio ($df = 1$ and $N_1 + N_2 - 6$) required for the α level of significance; $(P + Q)$ is a function of (a) the aptitude value in question, (b) the number of subjects in Treatment 1 (N_1), (c) the number of subjects in Treatment 2 (N_2), (d) the mean and variance for the aptitude variable and the square of the aptitude value, (e) the correlation between the aptitude variable and the square of the aptitude variable, and (f) the correlations of the criterion with the aptitude variable and with the square of the aptitude variable; and S_e^2 is the error sum of squares from model [14]. The actual expression for $(P + Q)$ for the two-aptitude case can be found in Johnson and Jackson (1959, p. 443). If the square of the aptitude is treated as a second aptitude variable, then this expression is appropriate to the single-aptitude, curvilinear case.

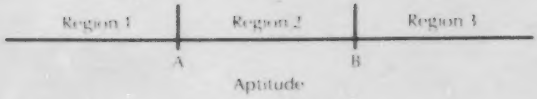
Bounding values for regions of significance are given by the following expression:

$$D^2 - \frac{F\alpha(P + Q)S_e^2}{N_1 + N_2 - 6} = 0 \,. \qquad [22]$$

Expression [22] is a fourth-order (quartic) equation involving the aptitude variable. ATICURV solves this quartic equation by first finding a solution for a resolvent cubic equation and then using this solution to obtain all four roots of the original quartic equation. Expression [22] yields two, four, or no real solutions corresponding to bounding values.

After calculating bounding values, program ATICURV determines the locations of regions of significance with regard to these bounding values. A bounding value is an aptitude score with a region of significance occurring on one side (above or below) of that score and a region of nonsignificance occurring on the other side of that score.

Consider the following example in which two bounding values (A and B) have been obtained.



In this example, either (a) Region 2 may involve significance while Regions 1 and 3 involve nonsignificance or (b) Region 2 may involve nonsignificance with Regions 1 and 3 involving significance. If the within-treatment regressions intersect, then it is easy to determine which case—(a) or (b)—is true. A region containing an intersection (point of zero difference between the regression lines) is always a region of nonsignificance. If Region 1 or Region 3 contains an intersection, then case (a) is true. On the other hand, if Region 2 contains an intersection, then case (b) is true.

Locating regions of significance for curvilinear regressions, however, is not always as simple as shown above. As demonstrated in Figure 17, two nonparallel curvilinear regressions need not intersect. Consider situation (a) in Figure 17. Two bounding values (A and B) exist in this situation. Does the region between A and B involve significance or nonsignificance? One might guess that the region between A and B is a region of nonsignificance, since the minimal distance between group regressions falls in this region. However, such a guess may prove false. The distances between regressions within a region of significance may actually be smaller than the distances within a bordering region of nonsignificance—such is the case when a single region of significance is found in the one-predictor, linear-regressions case (ATILIN1) or when an elliptical region is found in the two-predictor, linear-regressions case (ATILIN2). Without further information, it is impossible to determine if the region between A and B is one of significance or nonsignificance.

Situation (b) in Figure 17 also presents a problem. No bounding values are obtained, so there is only one region. However, does that region involve significance or nonsignificance? Without additional information it is impossible to tell. Program ATICURV deals with such

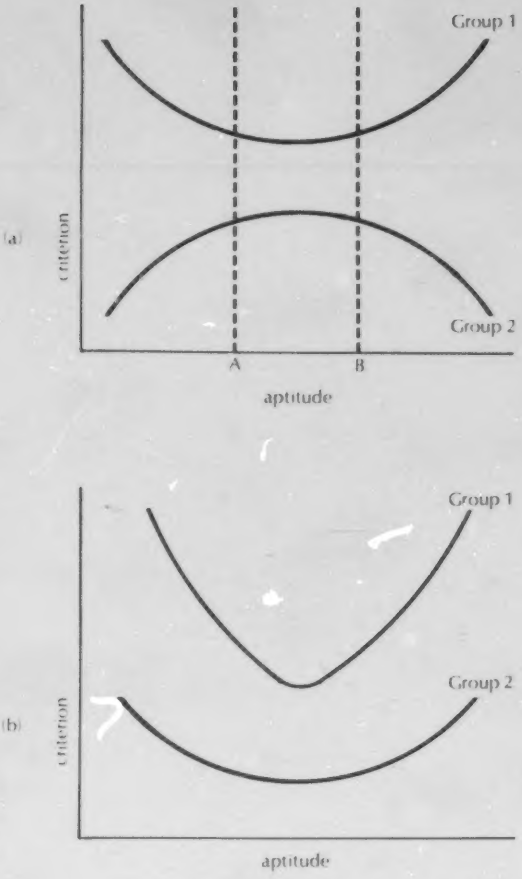


Figure 17. *Two examples of nonparallel and nonintersecting regressions. In situation (a), points A and B represent bounding values for regions of significance. Without additional information, it is impossible to determine where significance lies with respect to A and B. In situation (b), no bounding values are obtained. Without additional information, it is impossible to determine if Group 1 is significantly superior for all aptitude values or if there is no significant difference for any aptitude value.*

problems in the following manner. The midpoint of each region falling within the observed aptitude values is calculated. The significance of the difference (distance) between regressions is evaluated at each midpoint (if expression [21] is true for a midpoint, then the difference is significant at that point). A significant difference at a midpoint indicates that the corresponding region is a region of significance, while a nonsignificant difference at a midpoint indicates a region of nonsignificance. Program ATICURV reports the upper and lower bounds for each region of significance falling within the range of observed aptitude values.

Importance of the region of significance for curvilinear regressions. Recall considerations previously made with regard to the importance of a region of significance (p. 33). Such considerations also apply to regions of significance defined with regard to curvilinear regressions within treatments. The importance of a region of significance is (1) a positive function of the proportion of total observations that fall within that region and (2) a negative function of the amount of overlapping between the treatments within that region. Recall that a Treatment 1 observation evidences overlap if it falls closer to the Treatment 2 regression line than the Treatment 1 regression line. In other words, an observation is counted as overlapping if it falls on the "wrong" side of the midline between the group regression lines. The midline between curvilinear (quadratic) regression lines is given by the following equation:

$$Mpt(A_i) = [\overline{Y}_1 + b_{11}(A_i - \overline{A}_1) + b_{21}(A_i^2 - \overline{A}_1^2) + \overline{Y}_2 + b_{12}(A_i - \overline{A}_2) - b_{22}(A_i^2 - \overline{A}_2^2)] / 2$$

where $Mpt(A_i)$ is the midline criterion score for the aptitude variable equal to A_i; $\overline{Y}_1$ and $\overline{Y}_2$ are the criterion means for the two treatments; $\overline{A}_1$ and $\overline{A}_2$ are the aptitude means for the two treatments; and the b's are from [17] and [18].

References

Johnson, P. O., & Jackson, R. W. B. *Modern statistical methods.* Chicago: Rand-McNally, 1959.

Johnson, P. O., & Neyman, J. Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs*, 1936, *1*, 57-93.

Kelly, F. J., Beggs, D. L., McNeil, K. A., Eichelberger, T., & Lyon, J. *Research design in the behavioral sciences: Multiple regression approach.* Carbondale: Southern Illinois University Press, 1969.

Wunderlich, K. W., & Bori :h, G. D. Curvilinear extensions to Johnson-Neyman regions of significance and some applications to educational research. Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois, April, 1974.

CHAPTER VI

XGROUPS

Program Description

This program performs a treatment-by-blocks analysis of variance for two treatment groups and two levels of the aptitude variable. Within each treatment, high- and low-aptitude categories are formed by selecting extreme cases. The percentage of *Ss* in the sample included in the extreme aptitude categories (and, therefore, included in the treatment-by-blocks analysis of variance) is free to vary. If 10% of the *Ss* are to be included, then the 10% of the Treatment 1 *Ss* with the highest aptitude scores are assigned to the high-aptitude category and the 10% of the Treatment 1 *Ss* with the lowest aptitude scores are assigned to the low-aptitude category. In analogous fashion, 10% of the Treatment 2 *Ss* are assigned to each of the aptitude categories. Program XGROUPS performs multiple treatment-by-blocks analyses each corresponding to a different percentage of the sample to be included in the extreme group. Two options are available. The first option allows the user to input the exact percentages to be included in the extreme groups (a maximum of six percentages is allowed), and a separate treatment-by-blocks analysis is then computed for each percentage. If the second option is selected, then extreme groups with 10, 20, 30 and 40 percent of the sample are constructed and the corresponding four analyses are performed. Statistical power $(1 - \beta)$ is calculated for main and interaction effects in each analysis, allowing the user to determine the extreme group size resulting in the greatest statistical power. Program outputs sums of squares, degrees of freedom, mean squares, *F*-values, probabilities, power estimates and cell means for treatments, aptitude level (high vs. low), and treatments by levels. A flow chart for program XGROUPS is presented in Figure 18.

Figure 18. *Flow chart for program XGROUPS.*

Program Input

Card 1 alphanumeric title card Col 1-80

Card 2 parameter card

Col 1-5	*N* for Group 1 (maximum = 200)
Col 6-10	*N* for Group 2 (maximum = 200)
Col 12	missing data option 0 = all data valid 1 = blanks are invalid 2 = blanks and zeroes are invalid
Col 14	percent of sample option A zero in this column is a default value: the program will perform analyses with 10, 20, 30 and 40 percent of the sample in each of the extreme groups. An alternative set of percentages can be requested by indicating the number of percentages desired (maximum = 6) and then including the alternative percentages on Card 3 (see below).
Col 16	option for table of predictor and criterion scores listed by subject within treatments. If this option is taken, ID codes will be read according to format and will be printed out (A5) along with corresponding predictor and criterion scores for subjects in each treatment group. Subjects with missing data will not be listed in this table. 0 = no table 1 = list ID codes and scores (begin format cards with A mode field)
Col 18	no. of cards per subject in Group 1
Col 20	no. of cards per subject in Group 2

Card 3 (optional—include only if Col 14 is not 0) indicate integer percent levels desired in Col 1-3, 4-6, 7-9, etc., as necessary (maximum = 6)

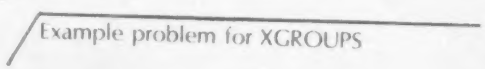
Card 4	format for Group 1 Col 1-80 followed by Group 1 data
Card 5	format for Group 2 Col 1-80 followed by Group 2 data
Card 6	blank (after last problem) for multiple problems repeat cards 1-5, omitting data

Data cards should contain subject ID codes (if desired), the aptitude score and then the criterion score. If Col 16 on the parameter card is 0, then formats must specify two F-mode fields—the first field for the aptitude and the second for the criterion. If Col 16 on the parameter card is 1, then formats must specify an initial A-mode field (A5 or less) for the ID code and then the two F-mode fields.

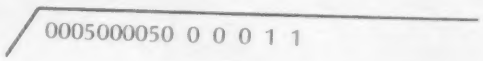
Example Problem

Data for this problem will be the first predictor and criterion given as sample data in Chapter VII (p.95) of this manual. Program control cards for this example problem are as follows.

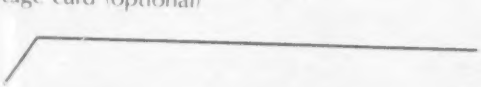
1. Alphanumeric title card



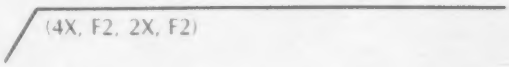
2. Parameter card



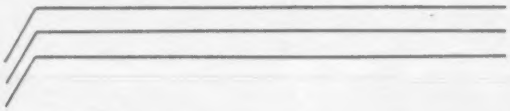
3. Cutoff percentage card (optional)



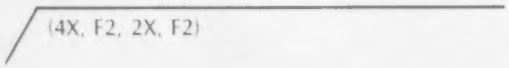
4. Format card for Group 1



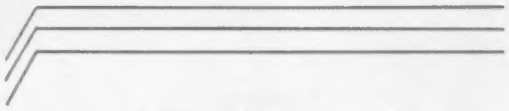
5. Group 1 data



6. Format card for Group 2

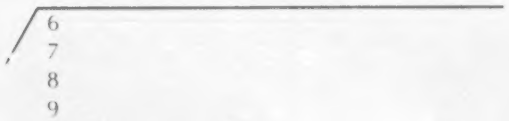


7. Group 2 data



8. Blank card (after last problem)

9. EOF



Printed output for this example problem is given in Figure 19.

```
EXTREME GROUPS ANALYSIS: LEVEL 1

NO. OF SUBJECTS IN TREATMENT 1 =  50
NO. OF SUBJECTS IN TREATMENT 2 =  50

REQUESTED CUTOFF PERCENTAGE = 10
ACTUAL CUTOFF PERCENTAGES:
     TREATMENT 1 =10.0000
     TREATMENT 2 =10.0000

EXTREME GROUP SIZES:
     NO. OF SUBJECTS IN EACH TREATMENT 1 GROUP =   5
     NO. OF SUBJECTS IN EACH TREATMENT 2 GROUP =   5


SOURCE          :         SS          DF          MS          F          P          POWER
GROUP(G)        :     125.00        1.00      125.00      1.9231     .181964     .265526
TREATMENT(T)    :     100.00        1.00      100.00      2.7692     .112375     .360000
GXT             :   17405.00        1.00    17405.00    267.7692     .000000    1.000000
ERROR           :    1040.00       16.00       65.00
TOTAL           :   18750.00       19.00


TREATMENT MEANS:
     TREATMENT 1 =   48.000
     TREATMENT 2 =   42.000

GROUP MEANS:
     HIGH =   47.500
     LOW  =   42.500

GROUP BY TREATMENT MEANS:
     TREATMENT 1 HI =   80.000
     TREATMENT 1 LO =   16.000
     TREATMENT 2 HI =   15.000
     TREATMENT 2 LO =   69.000
```

Figure 19. *Printed output for program XGROUPS example problem.*

```
EXTREME GROUPS ANALYSIS: LEVEL 2

NO. OF SUBJECTS IN TREATMENT 1 =  50
NO. OF SUBJECTS IN TREATMENT 2 =  50

REQUESTED CUTOFF PERCENTAGE = 20
ACTUAL CUTOFF PERCENTAGES:
     TREATMENT 1 =20.0000
     TREATMENT 2 =20.0000

EXTREME GROUP SIZES:
     NO. OF SUBJECTS IN EACH TREATMENT 1 GROUP =  10
     NO. OF SUBJECTS IN EACH TREATMENT 2 GROUP =  10


SOURCE          :        SS          DF          MS          F          P          POWER
GROUP(G)        :     90.00        1.00       90.00       .7286     .402826      .131991
TREATMENT(T)    :    160.00        1.00      160.00      1.2980     .261021      .204129
GXT             :  22562.50        1.00    22562.50    183.1454     .000000     1.000000
ERROR           :   4435.00       36.00      123.19
TOTAL           :  27247.50       39.00


TREATMENT MEANS:
     TREATMENT 1 =   49.750
     TREATMENT 2 =   45.750

GROUP MEANS:
     HIGH =   49.250
     LOW  =   46.250

GROUP BY TREATMENT MEANS:
     TREATMENT 1 HI =   75.000
     TREATMENT 1 LO =   24.500
     TREATMENT 2 HI =   23.500
     TREATMENT 2 LO =   68.000
```

Figure 19 (continued). *Printed output for program XGROUPS example problem.*

```
EXTREME GROUPS ANALYSIS: LEVEL 3

NO. OF SUBJECTS IN TREATMENT 1 =  50
NO. OF SUBJECTS IN TREATMENT 2 =  50

REQUESTED CUTOFF PERCENTAGE = 30
ACTUAL CUTOFF PERCENTAGES:
     TREATMENT 1 =30.0000
     TREATMENT 2 =30.0000

EXTREME GROUP SIZES:
     NO. OF SUBJECTS IN EACH TREATMENT 1 GROUP =  15
     NO. OF SUBJECTS IN EACH TREATMENT 2 GROUP =  15


SOURCE        :        SS          DF          MS          F          P          POWER
GROUP(G)      :       106.67       1.00       106.67       .7111     .407399     .131103
TREATMENT(T)  :       106.67       1.00       106.67       .7111     .407399     .131103
GXT           :     24401.67       1.00     24401.67    162.6778     .000000    1.000000
ERROR         :      8400.00      56.00       150.00
TOTAL         :     33015.00      59.00


TREATMENT MEANS:
     TREATMENT 1 =   49.833
     TREATMENT 2 =   47.167

GROUP MEANS:
     HIGH =   49.833
     LOW  =   47.167

GROUP BY TREATMENT MEANS:
     TREATMENT 1 HI =   71.333
     TREATMENT 1 LO =   28.333
     TREATMENT 2 HI =   28.333
     TREATMENT 2 LO =   66.000
```

Figure 19 (continued). *Printed output for program XGROUPS example problem.*

```
EXTREME GROUPS ANALYSIS: LEVEL 4

NO. OF SUBJECTS IN TREATMENT 1 =  50
NO. OF SUBJECTS IN TREATMENT 2 =  50

REQUESTED CUTOFF PERCENTAGE = 40
ACTUAL CUTOFF PERCENTAGES:
     TREATMENT 1 =40.0000
     TREATMENT 2 =40.0000

EXTREME GROUP SIZES:
     NO. OF SUBJECTS IN EACH TREATMENT 1 GROUP =  20
     NO. OF SUBJECTS IN EACH TREATMENT 2 GROUP =  20


SOURCE          :        SS          DF          MS          F          P          POWER
GROUP(G)        :      15.31        1.00       15.31        .0890      .763779    .052439
TREATMENT(T)    :      70.31        1.00       70.31        .4089      .531549    .092554
GXT             :   24325.31        1.00    24325.31     141.4614      .000000   1.000000
ERROR           :   13068.75       76.00      171.96
TOTAL           :   37479.69       79.00


TREATMENT MEANS:
     TREATMENT 1 =   49.375
     TREATMENT 2 =   47.500

GROUP MEANS:
     HIGH =   48.875
     LOW  =   48.000

GROUP BY TREATMENT MEANS:
     TREATMENT 1 HI =   67.250
     TREATMENT 1 LO =   31.500
     TREATMENT 2 HI =   30.500
     TREATMENT 2 LO =   64.500


END OF JOB.
```

Figure 19 (continued). *Printed output for program XGROUPS example problem*

1. General

The consideration of statistical power is crucial to any field of inquiry in which researchers consistently fail to reject the null hypothesis. This has tended to be the case in the field of aptitude-treatment interaction (ATI) research wherein relatively few significant interactions have been reported. For example, Bracht (1971), who conducted an extensive review of the ATI literature, could report finding only five significant interactions among 90 studies which hypothesized an aptitude-by-treatment interaction.

Even though seemingly crucial to the intrepetation of an ATI study that fails to reject the null hypothesis, statistical power is rarely, if ever, reported in ATI research. This circumstance is no doubt influenced in part by the complexity of the concept of power and sometimes by the laborious calculations that often need be performed in the absence of any handy programming routines.

2. Power

Cohen (1969) in his *Statistical Power Analyses for the Behavioral Sciences* provides tables that are reasonably good approximations for estimating levels of power for the analysis of variance. Cohen used the following three sources for constructing these tables:

A. Laubscher's (1960) square root normal approximation of noncentral *F*, given by the formula:

$$z_{1-\beta} = \frac{\left[2(u+\lambda) - \frac{u+2\lambda}{u+\lambda}\right]^{1/2} - \left[(2v-1)\frac{uF_c}{v}\right]^{1/2}}{\left[\frac{uF_c}{v} + \frac{u+2\lambda}{u+\lambda}\right]^{1/2}} \qquad [23]$$

where $z_{1-\beta}$ is a normal deviate which determines the value of power; u, the degrees of freedom in the numerator of the obtained F-ratio (F_O); v, the degrees of freedom in the denominator of F_O; $\lambda = F_O u$; and F_C, the F-ratio required for significance. The value of power (usually symbolized as $1 - \beta$) is the probability of obtaining a normal deviate at least as small as $z_{1-\beta}$. In other words, the value of power equals the normal curve area from $-\infty$ to $z_{1-\beta}$. Given that power is symbolized as $1 - \beta$, then β is the complement of power or alternatively β is the risk of making a Type II error, i.e., of accepting the null hypothesis when it is false. Note that λ and, therefore, power are functions of the cell frequency, n. As employed here, cell frequency refers to the number of scores upon which each group mean relevant to the F_O comparison is based. The value of F_O approximates $n\sigma_{\bar{X}}^2/\sigma^2$. Thus, F_O and λ both increase as cell frequency increases. The square root normal approximation [23] is best suited for generating power values when n and F_O are not small.

B. When n and F_O are small a second approximation is most appropriate. This is Laubscher's cube root normal approximation of noncentral F, given by the formula:

$$z_{1-\beta} = \frac{1 - \frac{2(u+2\lambda)}{9(u+\lambda)^2} - \left(1 - \frac{2}{9v}\right)\left(\frac{uF_C}{u+\lambda}\right)^{1/3}}{\left[\left(\frac{2}{9v}\right)\left(\frac{uF_C}{u+\lambda}\right)^{2/3} + \frac{2(u+2\lambda)}{9(u+\lambda)^2}\right]^{1/2}} \quad (24)$$

where u, v, F_C and λ are as defined above.

C. The final source of values for Cohen's tables were tables provided him by the National Bureau of Standards. These tables provide exact power values for combinations of a limited number of values for n, u, F_O and F_C. Cohen reconciled his approximations with the exact values that were available from the National Bureau of Standards.

Program XGROUPS incorporates a power estimating function (function POWER) which provides power estimates corresponding to Cohen's tabled values for differing values of v, F_O and F_C for the case

in which $u = 1$. These power values are generated through using both the square root and the cube root approximations, depending upon whether n is large or small. These approximations give normal deviate values corresponding to power (i.e., $z_{1-\beta}$). To obtain the probability value associated with this z, the following approximation derived from Hastings (1955) is used:

(a) the absolute value of the z score is taken;

(b) then the absolute value of the z score is used in the formula:

$$p = \frac{.5}{(1 + c_1 z + c_2 z^2 + c_3 z^3 + c_4 z^4)^4} \qquad [25]$$

where $c_1 = .916854$, $c_2 = .115194$, $c_3 = .000344$ and $c_4 = .091527$;

(c) then if the z score is positive, power is equal to $1 - p$ or, if the z score is negative, power is equal to p.

The reason for step (c) is that the above approximation for the probability associated with a z score gives the probability of obtaining a z score as extreme as that obtained and with the same sign as that obtained. In other words, equation [25] gives the smaller normal curve area bounded by the z score. With regard to power, however, the normal curve area from $-\infty$ to the z score is of interest. Equation [25] gives this area when the z score is negative. However, when the z score is positive, equation [25] gives the complement of the desired area. Therefore, step (c) is included to set power equal to p when the z score is negative and equal to the complement of p when the z score is positive.

3. *Accuracy of Function POWER*

Borich and Godbout (1974) tested the accuracy of the two approximations for power against Cohen's tabled values by performing 96 calculations of power with different values of n (cell frequency), F_O (the obtained F) and F_C (the F value needed for significance at the .05

level). The value of u, the number of degrees of freedom in the numerator, was always one, as interest was limited to effects involving a single degree of freedom. Therefore, this accuracy test was actually limited to certain combinations of values for F_O and n (F_C being determined by n). The square root approximation matched closest to Cohen's tabled values when n was 4 or greater, whereas the cube root approximation matched closest when n was less than 4. Function POWER, therefore, employs the square root approximation when $n \geq 4$ and the cube root approximation when $n < 4$.

4. *Cell Frequency*

A caution should be expressed concerning the meaning of cell frequency, n, as the present usage of the term is somewhat unorthodox. Cell frequency represents the number of scores producing a mean of interest. The means of interest are those upon which the obtained F-ratio (F_O) is based. As an example, one may consider a 2×2 design with 10 scores per cell. The F_O for the row effect is based upon two means each of which is determined by 20 scores. Therefore, the value of n is 20 for determination of the power of the significance test of the row effect. Similarly, the value of n is 20 for the determination of the power of the significance test of the column effect. In contrast, the row $\times$ column interaction is based upon comparison of all four cell means and the value of n is 10 for the determination of the power of the significance test of the interaction. It should be noted that *the cell frequency* (as presently defined) *is 20 for the test of the main effects but only 10 for the test of the interaction.* This lower cell frequency associated with the interaction illustrates an interesting general principle concerning power analyses. Power is a positive function of cell frequency. Given equal effect sizes and degrees of freedom, the lower cell frequency for the interaction test implies that the power of the test of the interaction will be lower than the power of the tests of the main effects. This lower power for interaction tests has been generally overlooked by the ATI researcher.

Statistical power is a positive function of the cell frequency, the effect size, and the alpha level chosen for significance. An increase in any one of these three quantities will cause an increase in power. As an illustration of the influence of the alpha level, consider an experimenter who has just calculated a 2×2 extreme groups analysis for which he has hypothesized that an interaction would be significant at the .01 level. The experimenter fails to reject the null hypothesis and estimates that the power of the analysis is .60; that is, he risks missing an interaction in four studies out of every ten. The experimenter is concerned that he may have chosen too low a value for alpha and that this low value alone may have unduly limited the power of his experiment and thus may have increased his chances of making a Type II error. But could he appreciably increase his power by raising the alpha level in future studies? If a desirable level of power, such as one between .70 and .80, could not be obtained even by increasing alpha to as high as .10, this researcher either should abandon this particular research or should consider a more powerful design.

Power is most usually raised in a given experiment by increasing the number of subjects in that experiment. If power is estimated and found to be low, the cost of improving it by increasing sample size will usually prove to be well worth the effort. If, on the other hand, power is estimated and found to be high, say, in the neighborhood of .80 to .90, the researcher will find that an increase in n, even of a large magnitude, may not substantially increase power further. It is not unusual for power values as great as .90, or higher, to demand sample sizes that exceed an experimenter's resources. The benefit of increasing sample size to gain power is therefore greatest for the researcher whose experiment has initially moderate or weak power.

Another way of increasing power is to increase the effect size. This is the strategy employed in the extreme groups analysis. The ATI researcher who chooses to employ the extreme groups technique is faced with a trade-off: What percentage of the distribution should he

assign to the two extremes in order to have a reasonable level of power? This question presents a dilemma which may be clarified by considering a practical example. Consider a sample of 100 subjects from which two extreme groups are to be selected for analysis. The extreme groups strategy dictates that the two groups be formed so that one includes the subjects scoring highest on the aptitude and the other includes the subjects scoring lowest on the aptitude. Each extreme group can contain as many as 50 subjects as an upper limit (the high group being subjects above the median and the low group being subjects below the median), or each extreme group can contain only a few subjects, e.g., five (the high group being the individuals with the five highest scores and the low group being the individuals with the five lowest scores). As the two groups become more extreme, the corresponding increment in the difference between their means (i.e., the effect size) causes power to increase. However, as the two groups become more extreme, the number of subjects in a group (i.e., the cell frequency) decreases and power diminishes. Thus, as the selected groups become more extreme, the presence of antagonistic effects on power leaves unclear what the change in power will be. Program XGROUPS has been designed to assist the ATI researcher in reconciling the need for increased power with a resultant decrease in sample size, i.e., to help him choose the number of cases that when assigned to extreme groups yields the most acceptable level of power.

6. *Treatment × Blocks ANOVA*

The standard analysis for the extreme groups design is a treatment × blocks analysis of variance. Program XGROUPS is appropriate to a 2 × 2 extreme groups design involving two treatments and two extreme aptitude groups within each treatment. An aptitude variable is used to establish high and low categories within each treatment. For Treatment 1, equal percentages of the Ss are assigned to high and low extreme groups. The same percentages of Treatment 2 Ss are

assigned to the two extreme groups.* If the total numbers of Treatment 1 and Treatment 2 Ss are not equal, then the resulting extreme groups design will involve unequal *n*'s. While unequal *n*'s often greatly complicate calculation and interpretation of an ANOVA, this is not true in the present case. Since Program XGROUPS forms equal-sized high and low aptitude groups within a given treatment, the cell frequencies in the resulting extreme groups design will always be proportional and conventional ANOVA calculations and interpretations are applicable (Kirk, 1968; Winer, 1971). Program XGROUPS, therefore, employs conventional ANOVA calculation techniques. Program output includes mean squares for treatments, levels (high vs. low), treatments × levels, and residual error. The *F*-ratio for treatments × levels is the test for aptitude-treatment interaction.

---

*Formation of extreme groups can be complicated by Ss with equal aptitude scores. Consider an aptitude that assumes integer values from 0 to 10. Say five Ss are to be included in the high category but that four Ss received a maximum score of 10 while three Ss received the next highest possible score of 9. Which five Ss should be included in the high category? In this case, program XGROUPS forms a high group consisting of the four Ss with scores of 10 and a single S (selected by a random number function) with a score of 9.

References

Borich, G. D., & Godbout, R. C. Extreme groups designs and the calculation of statistical power. *Educational and Psychological Measurement*, 1974, *34*, 663-675.

Bracht, G. H. Experimental factors related to aptitude-treatment interactions. *Review of Educational Research*, 1971, *40*, 627-645.

Cohen, J. *Statistical power analyses for the behavioral sciences*. New York: Academic Press, 1969.

Hastings, C., Jr. *Approximations for digital computers*. Princeton: Princeton University Press, 1955.

Kirk, R. E. *Experimental design: Procedures for the behavioral sciences*. Belmont, California: Brooks/Cole, 1968.

Laubscher, N. F. Normalizing the noncentral t and F distributions. *Annals of Mathematical Statistics*, 1960, *31*, 1105-1112.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971.

CHAPTER VII

SAMPLE DATA

TABLE OF PREDICTOR AND CRITERION SCORES LISTED BY SUBJECT WITHIN TREATMENTS

TREATMENT 1

ID	PREDICTOR 1	PREDICTOR 2	CRITERION
001	10.000	70.000	5.000
002	15.000	70.000	15.000
003	20.000	70.000	15.000
004	25.000	75.000	20.000
005	30.000	75.000	20.000
006	20.000	80.000	25.000
007	35.000	80.000	25.000
008	40.000	80.000	25.000
009	30.000	85.000	30.000
010	40.000	85.000	30.000
011	50.000	70.000	30.000
012	25.000	70.000	35.000
013	30.000	65.000	35.000
014	45.000	65.000	35.000
015	55.000	65.000	35.000
016	35.000	75.000	40.000
017	40.000	70.000	40.000
018	50.000	65.000	40.000
019	35.000	60.000	45.000
020	55.000	55.000	45.000
021	60.000	60.000	45.000
022	65.000	55.000	45.000
023	40.000	50.000	50.000
024	45.000	50.000	50.000
025	50.000	50.000	50.000
026	60.000	50.000	50.000
027	65.000	50.000	50.000
028	70.000	50.000	50.000
029	50.000	45.000	55.000
030	50.000	45.000	60.000
031	55.000	45.000	60.000
032	60.000	45.000	60.000
033	65.000	40.000	60.000
034	70.000	40.000	60.000
035	75.000	40.000	60.000
036	60.000	40.000	65.000
037	70.000	35.000	65.000
038	80.000	35.000	65.000
039	60.000	35.000	70.000
040	65.000	30.000	70.000
041	70.000	30.000	70.000
042	75.000	30.000	70.000
043	85.000	25.000	70.000
044	70.000	25.000	75.000
045	80.000	25.000	75.000
046	75.000	20.000	80.000
047	85.000	20.000	80.000
048	90.000	15.000	80.000
049	85.000	15.000	85.000
050	90.000	5.000	85.000

TREATMENT 2

ID	PREDICTOR 1	PREDICTOR 2	CRITERION
051	0.000	60.000	70.000
052	5.000	70.000	65.000
053	5.000	55.000	75.000
054	10.000	75.000	75.000
055	15.000	70.000	60.000
056	15.000	60.000	70.000
057	20.000	75.000	75.000
058	25.000	75.000	55.000
059	25.000	65.000	70.000
060	30.000	70.000	60.000
061	30.000	45.000	65.000
062	30.000	65.000	75.000
063	35.000	55.000	50.000
064	35.000	75.000	55.000
065	40.000	70.000	60.000
066	40.000	60.000	70.000
067	40.000	55.000	75.000
068	45.000	50.000	55.000
069	45.000	75.000	65.000
070	50.000	65.000	45.000
071	50.000	25.000	65.000
072	55.000	65.000	40.000
073	55.000	40.000	50.000
074	55.000	30.000	55.000
075	55.000	55.000	70.000
076	60.000	70.000	55.000
077	65.000	55.000	30.000
078	65.000	50.000	40.000
079	65.000	40.000	65.000
080	70.000	65.000	25.000
081	70.000	45.000	35.000
082	70.000	15.000	45.000
083	70.000	50.000	55.000
084	70.000	40.000	60.000
085	75.000	30.000	20.000
086	75.000	20.000	30.000
087	75.000	60.000	40.000
088	75.000	55.000	50.000
089	80.000	45.000	15.000
090	80.000	35.000	45.000
091	80.000	30.000	55.000
092	85.000	20.000	10.000
093	85.000	10.000	15.000
094	85.000	0.000	25.000
095	85.000	45.000	35.000
096	85.000	35.000	45.000
097	90.000	25.000	0.000
098	90.000	15.000	10.000
099	90.000	10.000	20.000
100	90.000	55.000	30.000

CHAPTER VIII

ANNOTATED BIBLIOGRAPHY

Abelson, R. P. A note on the Neyman-Johnson technique. *Psychometrika*, 1953, *18*(3), 213-218.

This article discusses the rationale for the Johnson-Neyman technique and the hypotheses that should be preliminarily tested before the technique is used. Specifically, the article suggests testing the hypotheses of (1) homogeneity of group variances and (2) equality of group regression slopes. If hypothesis (1) is rejected, it is theoretically not permissible to continue. If hypotheses (1) and (2) are both accepted, the author suggests testing the hypothesis that the intercepts are equal for the two groups. If hypothesis (1) is accepted and (2) is rejected, then the Johnson-Neyman technique is utilized. The author presents a method for the computation of the regions of significance for any number of predictors.

Cahen, L. S., & Linn, R. L. Regions of significant criterion differences in aptitude-treatment-interaction research. *American Educational Research Journal*, 1971, *8*(3), 521-530.

This article compares three techniques for determining regions of significant criterion differences when the effect of treatment interacts with the aptitude (predictor value) of the subjects in question. Compared are the Johnson-Neyman technique, the Potthoff modification of the Johnson-Neyman technique, and the Erlander and Gustavsson method. The authors demonstrate that the three techniques differ in their estimates of the size of the region of significance.

Johnson, P. O., & Fay, L. C. The Johnson-Neyman technique, its theory and application. *Psychometrika*, 1950, *15*, 349-367.

A detailed theoretical derivation of the Johnson-Neyman technique for determining regions of significance for interacting regression lines is presented. Also presented is a step-by-step computational ex-

ample problem in which there are two covariates (aptitudes), two treatments and one criterion measure.

Johnson, P. O., & Jackson, R. W. B. Special applications of multivariate analysis. In *Modern statistical methods*. Chicago: Rand McNally, 1959, 410-455.

The authors discuss the Johnson-Neyman method for comparing groups that have been measured on some aptitude (covariate, here called "matching" variable). The authors first consider using the technique when the aptitude is qualitatively or categorically defined (e.g., sex, race). Second, the authors discuss at length the more frequent case in which the aptitude variable is quantitatively defined (measurable across a range). Within this case, two examples are dealt with, one having a single aptitude and the second having two aptitudes.

Johnson, P. O., & Neyman, J. Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs*, 1936, *1*, 57-93.

This article contains the original formulations of the Johnson-Neyman technique for determining regions of significance. The greater part of the article is devoted to a detailed mathematical derivation of the technique, although space is also reserved for a discussion of the kinds of research problems for which the technique is appropriate. A numerical example is presented.

Koenker, R. H., & Hansen, C. W. Steps for the application of the Johnson-Neyman technique—A sample analysis. *Journal of Experimental Education*, 1942, *16*(3), 164-173.

A computational example using the Johnson-Neyman technique is presented. The problem analysis has two groups, one criterion, and two aptitude variables. A detailed computational procedure is presented.

Potthoff, R. F. On the Johnson-Neyman technique and some extensions thereof. *Psychometrika*, 1964, *29*, 241-255.

The article starts by reviewing the Johnson-Neyman technique, suggesting when it should and should not be used. Several modifications of the technique are then presented which:

(1) consider regions of significance as confidence intervals;
(2) use simultaneous confidence intervals instead of plotting the region of significance;
(3) consider the case having more than two groups;
(4) consider the case having more than one criterion.

Several simple numerical examples are presented.

Walker, H., & Lev., J. Analysis of covariance. In *Statistical inference*. New York: Holt, Rinehart and Winston, 1953, 387-412.

This chapter presents a derivation of the analysis of covariance model with a computational example having one covariate (aptitude), two treatments (groups, populations) and one criterion. Also presented are (1) an *F*-test for the hypothesis of equality of group regression lines, (2) an *F*-test for the hypothesis of equality among adjusted group (criterion) means, and (3) an *F*-test for the hypothesis of linearity of the regression line based on group means.

For an example having one covariate (aptitude), two treatments (groups, populations) and one criterion, a method for determining (1) the point of nonsignificance (the intersection of group regression lines), (2) the region of nonsignificance (those covariate values for which the treatments do not differ significantly), and (3) the regions of significance (those covariate values for which the treatments differ significantly) is presented.

Finally, for an example having two covariates, a method for determining the regions of significance and nonsignificance is presented. These regions are defined by all combinations of covariate values for which differences in the treatments do and do not differ significantly.